

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: May 23, 2003, 05:34:57 ; Search time 2038 Seconds  
(without alignments)  
628.324 Million cell updates/sec

Title: SEQ1-4EDITS  
Perfect score: 197  
Sequence: 1 ANSFLXXLRGSLXRCIXX.....XXAKXIFedVDLTAFWSKH 44

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n model -DEV=xlh  
-O=/cgn2.1/USPRO/SCHNITZER91/rnucat\_16052003\_160654\_26019/app\_query.fasta.1.199  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=SCHNITZER91@cgn.1.1.2496.@runat.16052003.160654.26019 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGQUEURY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMCOUT=120  
-WARN\_TIMCOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: gb\_da:\*  
16: gb\_fun:\*  
17: gb\_hum:\*  
18: gb\_in:\*  
19: gb\_mu:\*  
20: gb\_om:\*  
21: gb\_or:\*  
22: gb\_ov:\*  
23: gb\_pat:\*  
24: gb\_ph:\*  
25: gb\_pl:\*  
26: gb\_ro:\*  
27: gb\_sts:\*  
28: gb\_un:\*

29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	179	90.9	1386	6	AX207787 Sequence
2	179	90.9	1386	6	AX207788 Sequence
3	174	88.3	1386	6	AX207785 Sequence
4	174	88.3	1386	6	AX207786 Sequence
5	174	88.3	1386	6	AX212332 Sequence
6	174	88.3	1386	6	AX212333 Sequence
7	168	85.3	1386	6	AX149646 Sequence
8	167	84.8	1386	6	AX149645 Sequence
9	167	84.8	1386	6	AX212334 Sequence
10	162	82.2	1386	6	AX149644 Sequence
11	160	81.2	1257	6	AX422736 Sequence
12	160	81.2	1260	6	AX044041 Sequence
13	160	81.2	1260	6	AX149639 Sequence
14	160	81.2	1260	6	AX207783 Sequence
15	160	81.2	1260	6	AX212330 Sequence
16	160	81.2	1260	6	I00579 Sequence
17	160	81.2	1383	6	AX422734 Sequence
18	160	81.2	1386	6	AR070468 Sequence
19	160	81.2	1386	6	AX044042 Sequence
20	160	81.2	1386	6	AX044043 Sequence
21	160	81.2	1386	6	AX044044 Sequence
22	160	81.2	1386	6	AX044045 Sequence
23	160	81.2	1386	6	AX149640 Sequence
24	160	81.2	1386	6	AX207784 Sequence
25	160	81.2	1386	6	AX212331 Sequence
26	160	81.2	1386	6	E01708 Sequence
27	160	81.2	1386	6	I06643 Sequence
28	160	81.2	1386	6	I08112 Sequence
29	160	81.2	1389	6	E02492 Sequence
30	160	81.2	1756	6	I05477 Sequence
31	160	81.2	1792	9	BC034377 Sequence
32	160	81.2	1843	9	AX411026 Sequence
33	160	81.2	1843	9	HSPROT Sequence
34	157	79.7	1386	6	AX149643 Sequence
35	156	79.2	1386	6	AX149642 Sequence
36	156	79.2	1759	6	E01189 Sequence
37	155	78.7	1260	6	E00961 Sequence
38	155	78.7	1386	6	AX149641 Sequence
39	140	71.1	1499	10	MUSCP Sequence
40	140	71.1	1603	10	BC013896 Sequence
41	139	70.6	1543	6	AX401899 Sequence
42	139	70.6	1543	10	RNPPOC Sequence
43	138	70.1	1558	4	OC049933 Sequence
44	135	68.5	271	6	E02245 Sequence
45	135	68.5	1383	6	E02246 DNA sequence

RESULT 1

## ALIGNMENTS

AX207787  
LOCUS AX207787 1386 bp DNA linear PAT 31-AUG-2001  
DEFINITION Sequence 11 from Patent W00157193.  
ACCESSION AX207787  
VERSION AX207787.1 GI:15422463  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
1 (bases 1 to 1386)  
Gerlitz, B.E. and Jones, B.E.  
Protein c derivatives  
Patent: WO 0157193-A 11 09-AUG-2001;  
JOURNAL ELI LILLY AND COMPANY (US)  
FEATURES  
source  
1..1386  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 286 a 416 c 443 g 241 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 6.35e-24 Length: 1386  
Score: 179.00 Matches: 35  
Percent Similarity: 79.55% Conservative: 0  
Best Local Similarity: 79.55% Mismatches: 9  
Query Match: 90.86% Indels: 0  
Gaps: 0  
DB: 6  
SEQ1-4EDITS (1-44) x AX207787 (1-1386)  
QY 1 AlaasSerPheleu\*\*\*\*\*LeuArgGlnGlySerleu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
|||||  
Db 127 GCCACTCCTCTCTCGAGAGAGCTCGCTCAAGAGAGCGCTGAGCGGAGTGCATAGAGAG 186  
QY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspPThrLeuAlaPhe 40  
|||||  
Db 187 ATCTGTGACTTCGAGAGAGCGCAAGAAATTTTCGAGATGTGATGACACACTGGCCTTC 246  
QY 41 TrpSerLyshis 44  
|||||  
Db 247 TGGTCCAAGCAC 258  
RESULT 2  
AX207788 1386 bp DNA linear PAT 31-AUG-2001  
LOCUS AX207788  
DEFINITION Sequence 12 from Patent W00157193.  
ACCESSION AX207788  
VERSION AX207788.1 GI:15422464  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
1 (bases 1 to 1386)  
Gerlitz, B.E. and Jones, B.E.  
Protein c derivatives  
Patent: WO 0157193-A 12 09-AUG-2001;  
JOURNAL ELI LILLY AND COMPANY (US)  
FEATURES  
source  
1..1386  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 286 a 415 c 444 g 241 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 6.35e-24 Length: 1386  
Score: 179.00 Matches: 35  
Percent Similarity: 79.55% Conservative: 0  
Best Local Similarity: 79.55% Mismatches: 9  
Query Match: 90.86% Indels: 0

DB: 6 Gaps: 0  
SEQ1-4EDITS (1-44) x AX207788 (1-1386)  
QY 1 AlaasSerPheleu\*\*\*\*\*LeuArgGlnGlySerleu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
|||||  
Db 127 GCCACTCCTCTCTCGAGAGAGCTCGCTCAAGAGAGCGCTGAGCGGAGTGCATAGAGAG 186  
QY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspPThrLeuAlaPhe 40  
|||||  
Db 187 ATCTGTGACTTCGAGAGAGCGCAAGAAATTTTCGAGATGTGATGACACACTGGCCTTC 246  
QY 41 TrpSerLyshis 44  
|||||  
Db 247 TGGTCCAAGCAC 258  
RESULT 3  
AX207785 1386 bp DNA linear PAT 31-AUG-2001  
LOCUS AX207785  
DEFINITION Sequence 9 from Patent W00157193.  
ACCESSION AX207785  
VERSION AX207785.1 GI:15422461  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
1 (bases 1 to 1386)  
Gerlitz, B.E. and Jones, B.E.  
Protein c derivatives  
Patent: WO 0157193-A 9 09-AUG-2001;  
JOURNAL ELI LILLY AND COMPANY (US)  
FEATURES  
source  
1..1386  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 285 a 417 c 443 g 241 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 5.61e-23 Length: 1386  
Score: 174.00 Matches: 34  
Percent Similarity: 77.27% Conservative: 0  
Best Local Similarity: 77.27% Mismatches: 10  
Query Match: 88.32% Indels: 0  
Gaps: 0  
DB: 6  
SEQ1-4EDITS (1-44) x AX207785 (1-1386)  
QY 1 AlaasSerPheleu\*\*\*\*\*LeuArgGlnGlySerleu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
|||||  
Db 127 GCCACTCCTCTCTCGAGAGAGCTCGCTCAAGAGAGCGCTGAGCGGAGTGCATAGAGAG 186  
QY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspPThrLeuAlaPhe 40  
|||||  
Db 187 ATCTGTGACTTCGAGAGAGCGCAAGAAATTTTCGAGATGTGATGACACACTGGCCTTC 246  
QY 41 TrpSerLyshis 44  
|||||  
Db 247 TGGTCCAAGCAC 258  
RESULT 4  
AX207786 1386 bp DNA linear PAT 31-AUG-2001  
LOCUS AX207786  
DEFINITION Sequence 10 from Patent W00157193.  
ACCESSION AX207786  
VERSION AX207786.1 GI:15422462  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
1 (bases 1 to 1386)

AUTHORS Gerlitz, B.E. and Jones, B.E.  
TITLE Protein c derivatives  
JOURNAL Patent: WO 0157193-A 10 09-AUG-2001;  
ELI LILLY AND COMPANY (US)  
FEATURES Location/Qualifiers  
SOURCE 1. 1386  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 285 a 416 c 444 g 241 t  
ORIGIN

Alignment Scores:  
Pred. No.: 5.61e-23 Length: 1386  
Score: 174.00 Matches: 34  
Percent Similarity: 77.27% Conservative: 0  
Best Local Similarity: 77.27% Mismatches: 10  
Query Match: 88.32% Indels: 0  
Gaps: 0

SEQ1-4EDITS (1-44) x AX207786 (1-1386)

OY 1 AlaasnerPheLeu\*\*\*\*\*LeuArgGInGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
DB 127 GCCAAGCTCCTTCTCGAGAGAGCTCCGTCACGGAGCGAGCTGGAGCGAGTGCATAGAGAG 186

OY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
DB 187 ATCTGTGACTTCGAGAGAGCCAAAGAAATTTTGAAGATGTGATGACACACTGGCCTTC 246

OY 41 TrpSerLysHis 44  
DB 247 TGGTCCAAAGCAC 258

RESULT 5  
AX212332  
LOCUS AX212332 1386 bp DNA linear PAT 06-SEP-2001  
DEFINITION Sequence 8 from Patent WO0159084.  
ACCESSION AX212332.1 GI:15524088  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE Gerlitz, B.E., Grinnell, B.W. and Jones, B.E.  
AUTHORS  
TITLE Protein c derivatives  
JOURNAL Patent: WO 0159084-A 8 16-AUG-2001;  
ELI LILLY AND COMPANY (US)  
FEATURES Location/Qualifiers  
SOURCE 1. 1386  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 285 a 418 c 442 g 241 t  
ORIGIN

Alignment Scores:  
Pred. No.: 5.61e-23 Length: 1386  
Score: 174.00 Matches: 34  
Percent Similarity: 77.27% Conservative: 0  
Best Local Similarity: 77.27% Mismatches: 10  
Query Match: 88.32% Indels: 0  
Gaps: 0

SEQ1-4EDITS (1-44) x AX212332 (1-1386)

OY 1 AlaasnerPheLeu\*\*\*\*\*LeuArgGInGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
DB 127 GCCAAGCTCCTTCTCGAGAGAGCTCCGTCACGGAGCGAGCTGGAGCGAGTGCATAGAGAG 186

OY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
DB 187 ATCTGTGACTTCGAGAGAGCCAAAGAAATTTTGAAGATGTGATGACACACTGGCCTTC 246

OY 41 TrpSerLysHis 44  
DB 247 TGGTCCAAAGCAC 258

RESULT 5  
AX212332  
LOCUS AX212332 1386 bp DNA linear PAT 06-SEP-2001  
DEFINITION Sequence 8 from Patent WO0159084.  
ACCESSION AX212332.1 GI:15524088  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE Gerlitz, B.E., Grinnell, B.W. and Jones, B.E.  
AUTHORS  
TITLE Protein c derivatives  
JOURNAL Patent: WO 0159084-A 8 16-AUG-2001;  
ELI LILLY AND COMPANY (US)  
FEATURES Location/Qualifiers  
SOURCE 1. 1386  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 285 a 418 c 442 g 241 t  
ORIGIN

OY 41 TrpSerLysHis 44  
DB 247 TGGTCCAAAGCAC 258

RESULT 6  
AX212333  
LOCUS AX212333 1386 bp DNA linear PAT 06-SEP-2001  
DEFINITION Sequence 9 from Patent WO0159084.  
ACCESSION AX212333  
VERSION AX212333.1 GI:15524089  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE Gerlitz, B.E., Grinnell, B.W. and Jones, B.E.  
AUTHORS  
TITLE Protein c derivatives  
JOURNAL Patent: WO 0159084-A 9 16-AUG-2001;  
ELI LILLY AND COMPANY (US)  
FEATURES Location/Qualifiers  
SOURCE 1. 1386  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 285 a 417 c 443 g 241 t  
ORIGIN

Alignment Scores:  
Pred. No.: 5.61e-23 Length: 1386  
Score: 174.00 Matches: 34  
Percent Similarity: 77.27% Conservative: 0  
Best Local Similarity: 77.27% Mismatches: 10  
Query Match: 88.32% Indels: 0  
Gaps: 0

SEQ1-4EDITS (1-44) x AX212333 (1-1386)

OY 1 AlaasnerPheLeu\*\*\*\*\*LeuArgGInGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
DB 127 GCCAAGCTCCTTCTCGAGAGAGCTCCGTCACGGAGCGAGCTGGAGCGAGTGCATAGAGAG 186

OY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
DB 187 ATCTGTGACTTCGAGAGAGCCAAAGAAATTTTGAAGATGTGATGACACACTGGCCTTC 246

OY 41 TrpSerLysHis 44  
DB 247 TGGTCCAAAGCAC 258

RESULT 7  
AX149646  
LOCUS AX149646 1386 bp DNA linear PAT 08-JUN-2001  
DEFINITION Sequence 16 from Patent WO0136462.  
ACCESSION AX149646  
VERSION AX149646.1 GI:14348045  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE Gerlitz, B.E., Grinnell, B.W., Huang, L. and Jones, B.E.  
AUTHORS  
TITLE Protein c derivatives  
JOURNAL Patent: WO 0136462-A 16 25-MAY-2001;  
ELI LILLY AND COMPANY (US)  
FEATURES Location/Qualifiers  
SOURCE 1. 1386  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 288 a 416 c 442 g 240 t  
ORIGIN

Alignment Scores:

Pred. No.:	7.67e-22	Length:	1386
Score:	168.00	Matches:	32
Percent Similarity:	79.55%	Conservative:	3
Best Local Similarity:	72.73%	Mismatches:	9
Query Match:	85.28%	Indels:	0
DB:	6	Gaps:	0

SEQ1-4EDITS (1-44) x AX149646 (1-1386)

QY 1 AlaasSerPheLeu\*\*\*\*\*LeuArgGinglySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 DB 127 GCCAACTCCTTCCTGAGAGAGCTCCGCTCAAGGAGGAGCGGAGGATGATAGAGAG 186

QY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
 DB 187 ATCTGTGACTTGGAGAGGCCAAGGAATTTTCCAAATGTGATGACACACTGGCCTTC 246

QY 41 TrpSerLyshis 44  
 DB 247 TGGTCCACAGCAC 258

RESULT 8  
 AX149645 1386 bp DNA linear PAT 08-JUN-2001

LOCUS AX149645  
 DEFINITION Sequence 15 from Patent W00136462.  
 ACCESSION AX149645  
 VERSION AX149645.1 GI:14348044  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 Gerlitz, B.E., Grinnell, B.W., Huang, L. and Jones, B.E.  
 Protein c derivatives  
 Patent: WO 0136462-A 15 25-MAY-2001;  
 JOURNAL ELI LILLY AND COMPANY (US)  
 FEATURES  
 Location/Qualifiers  
 source 1..1386  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 289 a 415 c 442 g 240 t

ORIGIN

Alignment Scores:

Pred. No.:	1.19e-21	Length:	1386
Score:	167.00	Matches:	32
Percent Similarity:	77.27%	Conservative:	2
Best Local Similarity:	72.73%	Mismatches:	10
Query Match:	84.77%	Indels:	0
DB:	6	Gaps:	0

SEQ1-4EDITS (1-44) x AX149645 (1-1386)

QY 1 AlaasSerPheLeu\*\*\*\*\*LeuArgGinglySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 DB 127 GCCAACTCCTTCCTGAGAGAGCTCCGCTCAAGGAGGAGCGGAGGATGATAGAGAG 186

QY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
 DB 187 ATCTGTGACTTGGAGAGGCCAAGGAATTTTCCAAATGTGATGACACACTGGCCTTC 246

QY 41 TrpSerLyshis 44  
 DB 247 TGGTCCACAGCAC 258

RESULT 9  
 AX212334 1386 bp DNA linear PAT 06-SEP-2001

LOCUS AX212334  
 DEFINITION Sequence 10 from Patent W00159084.  
 ACCESSION AX212334  
 VERSION AX212334.1 GI:15524090

KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 (bases 1 to 1386)  
 Gerlitz, B.E., Grinnell, B.W. and Jones, B.E.  
 Protein c derivatives  
 Patent: WO 0159084-A 10 16-AUG-2001;  
 JOURNAL ELI LILLY AND COMPANY (US)  
 FEATURES  
 Location/Qualifiers  
 source 1..1386  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 288 a 416 c 442 g 240 t

ORIGIN

Alignment Scores:

Pred. No.:	1.19e-21	Length:	1386
Score:	167.00	Matches:	32
Percent Similarity:	77.27%	Conservative:	2
Best Local Similarity:	72.73%	Mismatches:	10
Query Match:	84.77%	Indels:	0
DB:	6	Gaps:	0

SEQ1-4EDITS (1-44) x AX212334 (1-1386)

QY 1 AlaasSerPheLeu\*\*\*\*\*LeuArgGinglySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 DB 127 GCCAACTCCTTCCTGAGAGAGCTCCGCTCAAGGAGGAGCGGAGGATGATAGAGAG 186

QY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
 DB 187 ATCTGTGACTTGGAGAGGCCAAGGAATTTTCCAAATGTGATGACACACTGGCCTTC 246

QY 41 TrpSerLyshis 44  
 DB 247 TGGTCCACAGCAC 258

RESULT 10  
 AX149644 1386 bp DNA linear PAT 08-JUN-2001

LOCUS AX149644  
 DEFINITION Sequence 14 from Patent W00136462.  
 ACCESSION AX149644  
 VERSION AX149644.1 GI:14348043  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 (bases 1 to 1386)  
 Gerlitz, B.E., Grinnell, B.W., Huang, L. and Jones, B.E.  
 Protein c derivatives  
 Patent: WO 0136462-A 14 25-MAY-2001;  
 JOURNAL ELI LILLY AND COMPANY (US)  
 FEATURES  
 Location/Qualifiers  
 source 1..1386  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"

BASE COUNT 288 a 417 c 441 g 240 t

ORIGIN

Alignment Scores:

Pred. No.:	1.05e-20	Length:	1386
Score:	162.00	Matches:	31
Percent Similarity:	75.00%	Conservative:	2
Best Local Similarity:	70.45%	Mismatches:	11
Query Match:	82.23%	Indels:	0
DB:	6	Gaps:	0

SEQ1-4EDITS (1-44) x AX149644 (1-1386)

QY 1 AlaasSerPheLeu\*\*\*\*\*LeuArgGinglySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20

```

Db 127 GCGAAGCTCTTCTCGAGAGAGCTCGTCACGCGAGCGGAGAGTCATAGAGAG 186
Oy 21 IlecysAspPhe*****AlaLys**IlephGluAspValAspAspThrLeuAlaPhe 40
Db 187 ATCTGTGACTTGGAGAGAGCCCAAGAAATTTTCCAAATGTGATGACACACTGGCCTTC 246
Oy 41 TrpSerLysHis 44
Db 247 TGGTCCCAAGCAC 258

RESULT 11
AX427736 1257 bp DNA linear PAT 20-JUN-2002
LOCUS DEFINITION Sequence 3 from Patent WO0232461.
AX427736
VERSION AX427736.1 GI:21537843
KEYWORDS
ORGANISM human.
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Andersen K.V., Freskgaard P.O. and Pedersen A.H.
TITLE Protein C or activated protein C-like molecules
JOURNAL APS (DK): MAXGEN HOLDINGS LTD (US)
MAXGEN
FEATURES
source
1..1257
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..>1257
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD35980.1"
/db_xref="GI:21537844"
/translation="ANSFLEELRHSSLEECIEICDEFEAKELFQNVNDTLAEWSKH
VDGQOCVLPLEHPGASLCGHGTCIDIGSFSCDRSGWGCGRFCQREVSFLNCLSDN
GGCTHYCLEEYGMRCSCAPGYKLGDLLQCHPAVKFPGCPMKMEKKRSLKRDTE
DOEDVDPRLLDGMTRGDSPMOYVLDSGAVLLHPSVVLTAACMDSEK
LLVRIGEDLARMEKELDLIDKEFYFNPNSKSTDDIALHLAOPATISQTTVP
CLPDSGLERLNOAGQETLVGTGSHSSRKEAKRNTFVLFKIPVPHNECEV
MSNVSENMCLAGLGDQDACEGDSGPMVASFPGTWELGLVSWGCGLLHNYGV
YTKVSRYLDMTHGHIROKAPQKSWAP"

BASE COUNT 264 a 375 c 406 g 212 t
ORIGIN

Alignment Scores:
Pred. No.: 2,266-20 Length: 1257
Score: 160.00 Matches: 31
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 70.45% Mismatches: 11
Query Match: 81.22% Indels: 0
DB: Gaps: 0

SEQ1-4EDITS (1-44) x AX427736 (1-1257)
Oy 1 AlaAsnSerPheLeu*****LeuArgGlnGlySerLeu***Arg***CysIle***** 20
Db 1 GCGAAGCTCTTCTCGAGAGAGCTCGTCACAGCGAGCGGAGAGTCATAGAGAG 60
Oy 21 IlecysAspPhe*****AlaLys**IlephGluAspValAspAspThrLeuAlaPhe 40
Db 61 ATCTGTGACTTGGAGAGAGCCCAAGAAATTTTCCAAATGTGATGACACACTGGCCTTC 120
Oy 41 TrpSerLysHis 44
Db 121 TGGTCCCAAGCAC 132

RESULT 12
AX044041 1260 bp DNA linear PAT 24-NOV-2000
LOCUS DEFINITION Sequence 6 from Patent WO0066754.

```

```

ACCESSION AX044041
VERSION AX044041.1 GI:11342920
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1260)
AUTHORS Gerlitz B.E. and Jones B.E.
TITLE Protein C derivatives
JOURNAL Patent: WO 0066754-A 6 09-NOV-2000;
ELI LILLY AND COMPANY (US)
FEATURES
source
1..1260
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..>1260
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD35980.1"
/db_xref="GI:21537844"
/translation="ANSFLEELRHSSLEECIEICDEFEAKELFQNVNDTLAEWSKH
VDGQOCVLPLEHPGASLCGHGTCIDIGSFSCDRSGWGCGRFCQREVSFLNCLSDN
GGCTHYCLEEYGMRCSCAPGYKLGDLLQCHPAVKFPGCPMKMEKKRSLKRDTE
DOEDVDPRLLDGMTRGDSPMOYVLDSGAVLLHPSVVLTAACMDSEK
LLVRIGEDLARMEKELDLIDKEFYFNPNSKSTDDIALHLAOPATISQTTVP
CLPDSGLERLNOAGQETLVGTGSHSSRKEAKRNTFVLFKIPVPHNECEV
MSNVSENMCLAGLGDQDACEGDSGPMVASFPGTWELGLVSWGCGLLHNYGV
YTKVSRYLDMTHGHIROKAPQKSWAP"

BASE COUNT 265 a 375 c 407 g 213 t
ORIGIN

Alignment Scores:
Pred. No.: 2,266-20 Length: 1260
Score: 160.00 Matches: 31
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 70.45% Mismatches: 11
Query Match: 81.22% Indels: 0
DB: Gaps: 0

SEQ1-4EDITS (1-44) x AX044041 (1-1260)
Oy 1 AlaAsnSerPheLeu*****LeuArgGlnGlySerLeu***Arg***CysIle***** 20
Db 1 GCGAAGCTCTTCTCGAGAGAGCTCGTCACAGCGAGCGGAGAGTCATAGAGAG 60
Oy 21 IlecysAspPhe*****AlaLys**IlephGluAspValAspAspThrLeuAlaPhe 40
Db 61 ATCTGTGACTTGGAGAGAGCCCAAGAAATTTTCCAAATGTGATGACACACTGGCCTTC 120
Oy 41 TrpSerLysHis 44
Db 121 TGGTCCCAAGCAC 132

RESULT 13
AX149639 1260 bp DNA linear PAT 08-JUN-2001
LOCUS DEFINITION Sequence 9 from Patent WO0136462.
AX149639
VERSION AX149639.1 GI:14348038
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1260)
AUTHORS Gerlitz B.E., Grinnell B.W., Huang L. and Jones B.E.
TITLE Protein C derivatives
JOURNAL Patent: WO 0136462-A 9 25-MAY-2001;
ELI LILLY AND COMPANY (US)
FEATURES
source
1..1260
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
<1..>1260
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD35980.1"
/db_xref="GI:21537844"
/translation="ANSFLEELRHSSLEECIEICDEFEAKELFQNVNDTLAEWSKH
VDGQOCVLPLEHPGASLCGHGTCIDIGSFSCDRSGWGCGRFCQREVSFLNCLSDN
GGCTHYCLEEYGMRCSCAPGYKLGDLLQCHPAVKFPGCPMKMEKKRSLKRDTE
DOEDVDPRLLDGMTRGDSPMOYVLDSGAVLLHPSVVLTAACMDSEK
LLVRIGEDLARMEKELDLIDKEFYFNPNSKSTDDIALHLAOPATISQTTVP
CLPDSGLERLNOAGQETLVGTGSHSSRKEAKRNTFVLFKIPVPHNECEV
MSNVSENMCLAGLGDQDACEGDSGPMVASFPGTWELGLVSWGCGLLHNYGV
YTKVSRYLDMTHGHIROKAPQKSWAP"

BASE COUNT 265 a 375 c 407 g 213 t
ORIGIN

Alignment Scores:
Pred. No.: 2,266-20 Length: 1260
Score: 160.00 Matches: 31
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 70.45% Mismatches: 11
Query Match: 81.22% Indels: 0
DB: Gaps: 0

SEQ1-4EDITS (1-44) x AX149639 (1-1260)

```

```

QY 1 AlaAsnSerPheLeu*****LeuArgGlnGlySerLeu***Arg***CysIle***** 20
    |||||
Db 1 GCCAACTCCTCTCGAGAGAGCTCCGCTCACAGCAGCCTGGAGGGGAGTGCATAGAGAG 60

QY 21 IleCysAspPhe*****AlaIys***IlePheGluAspValAspAspThrLeuAlaPhe 40
    |||||
Db 61 ATCTGTGACTTCGAGAGAGCCAAAGAAATTTTCCAAATGTGATGACACACTGGCCTTC 120

QY 41 TrpSerLysHis 44
    |||||
Db 121 TGGTCCAGACAC 132

RESULT 14
AX207783 1260 bp DNA linear PAT 31-AUG-2001
LOCUS AX207783
DEFINITION Sequence 7 from Patent WO0159084.
ACCESSION AX207783.1 GI:1542459
VERSION AX207783.1
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1260)
Gerlitz,B.E. and Jones,B.E.
TITLE
Protein c derivatives
PATENT: WO 0157193-A 7 09-AUG-2001.
JOURNAL
ELI LILLY AND COMPANY (US)
FEATURES
    source
        1. .1260
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT 265 a 375 c 407 g 213 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 2,26e-20 Length: 1260
Score: 160.00 Matches: 31
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 70.45% Mismatches: 11
Query Match: 81.22% Indels: 0
DB: 6 Gaps: 0

```

SEQ1-4EDITS (1-44) x AX207783 (1-1260)

```

QY 1 AlaAsnSerPheLeu*****LeuArgGlnGlySerLeu***Arg***CysIle***** 20
    |||||
Db 1 GCCAACTCCTCTCGAGAGAGCTCCGCTCACAGCAGCCTGGAGGGGAGTGCATAGAGAG 60

QY 21 IleCysAspPhe*****AlaIys***IlePheGluAspValAspAspThrLeuAlaPhe 40
    |||||
Db 61 ATCTGTGACTTCGAGAGAGCCAAAGAAATTTTCCAAATGTGATGACACACTGGCCTTC 120

QY 41 TrpSerLysHis 44
    |||||
Db 121 TGGTCCAGACAC 132

```

```

RESULT 15
AX212330 1260 bp DNA linear PAT 06-SEP-2001
LOCUS AX212330
DEFINITION Sequence 6 from Patent WO0159084.
ACCESSION AX212330
VERSION AX212330.1 GI:15524086
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 1260)
Gerlitz,B.E., Grinnell,B.W. and Jones,B.E.
TITLE
Protein c derivatives
PATENT: WO 0159084-A 6 16-AUG-2001;
JOURNAL

```

```

FEATURES
    source
        ELI LILLY AND COMPANY (US)
        Location/Qualifiers
        1. .1260
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
BASE COUNT 265 a 375 c 407 g 213 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 2,26e-20 Length: 1260
Score: 160.00 Matches: 31
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 70.45% Mismatches: 11
Query Match: 81.22% Indels: 0
DB: 6 Gaps: 0

```

SEQ1-4EDITS (1-44) x AX212330 (1-1260)

```

QY 1 AlaAsnSerPheLeu*****LeuArgGlnGlySerLeu***Arg***CysIle***** 20
    |||||
Db 1 GCCAACTCCTCTCGAGAGAGCTCCGCTCACAGCAGCCTGGAGGGGAGTGCATAGAGAG 60

QY 21 IleCysAspPhe*****AlaIys***IlePheGluAspValAspAspThrLeuAlaPhe 40
    |||||
Db 61 ATCTGTGACTTCGAGAGAGCCAAAGAAATTTTCCAAATGTGATGACACACTGGCCTTC 120

QY 41 TrpSerLysHis 44
    |||||
Db 121 TGGTCCAGACAC 132

```

Search completed: May 23, 2003, 06:44:56  
Job time : 2054 secs

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: May 23, 2003, 04:55:39 ; Search time 241 Seconds

(without alignments)  
411.153 Million cell updates/sec

Title: SEQ1-4EDITS

Perfect score: 197  
Sequence: 1 ANSLXLRGSLXKRCIX.....XXAKXIFEDVDTLAFWSKH 44

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=x1b  
-Q=gcgn2.1/USPTO/SCNITER511/unat\_16052003.160654.26010/app.query.fasta.1.199  
-DB=N.Geneseq\_101002 -QEXT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -DOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=SCNITER511.GCGN.1.1.263 -runat.16052003.160654.26010 -NCPU=6 -ICPU=3  
-NO\_XLPPX -NO\_MAP -LARGEDETER -NEG\_SCORES=0 -WAIT -LONGLOC -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YCAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq\_101002.\*  
1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	179	90.9	1386	22	AAD15227	Human protein C de
2	179	90.9	1386	22	AAD15228	Human protein C de
3	179	90.9	1386	22	AAH26365	DNA encoding human
4	179	90.9	1386	22	AAH26366	DNA encoding human
5	174	88.3	1386	22	AAD15225	Human protein C de
6	174	88.3	1386	22	AAD15226	Human protein C de
7	174	88.3	1386	22	AAH26363	DNA encoding human
8	174	88.3	1386	22	AAH26364	DNA encoding human
9	160	81.2	1245	21	AA246550	Truncated human pr
10	160	81.2	1257	24	ABR86039	Synthetic DNA enco
11	160	81.2	1260	7	AAH60001	Sequence encoding
12	160	81.2	1260	22	AAD15223	Human mature wild
13	160	81.2	1260	22	AAH26361	DNA encoding human
14	160	81.2	1260	22	AAH26361	DNA encoding human
15	160	81.2	1383	24	ABR86038	Synthetic DNA enco
16	160	81.2	1386	9	AAH81408	DNA encoding human
17	160	81.2	1386	10	AAH90024	Nascent human prot
18	160	81.2	1386	10	AAH90187	DNA encoding zymo
19	160	81.2	1386	18	AAH79724	Human protein C cd
20	160	81.2	1386	22	AAD15224	Human wild type pr
21	160	81.2	1386	22	AAH26362	DNA encoding human
22	160	81.2	1386	22	AAH26362	DNA encoding human
23	160	81.2	1386	22	AAH26362	DNA encoding human
24	160	81.2	1386	22	AAH26362	DNA encoding human
25	160	81.2	1386	22	AAH26362	DNA encoding human
26	160	81.2	1386	22	AAH26362	DNA encoding human
27	160	81.2	1386	22	AAH26362	DNA encoding human
28	160	81.2	1386	22	AAH26362	DNA encoding human
29	160	81.2	1386	22	AAH26362	DNA encoding human
30	160	81.2	1386	22	AAH26362	DNA encoding human
31	160	81.2	1386	22	AAH26362	DNA encoding human
32	160	81.2	1386	22	AAH26362	DNA encoding human
33	160	81.2	1386	22	AAH26362	DNA encoding human
34	160	81.2	1386	22	AAH26362	DNA encoding human
35	160	81.2	1386	22	AAH26362	DNA encoding human
36	160	81.2	1386	22	AAH26362	DNA encoding human
37	160	81.2	1386	22	AAH26362	DNA encoding human
38	160	81.2	1386	22	AAH26362	DNA encoding human
39	160	81.2	1386	22	AAH26362	DNA encoding human
40	160	81.2	1386	22	AAH26362	DNA encoding human
41	160	81.2	1386	22	AAH26362	DNA encoding human
42	160	81.2	1386	22	AAH26362	DNA encoding human
43	160	81.2	1386	22	AAH26362	DNA encoding human
44	160	81.2	1386	22	AAH26362	DNA encoding human
45	160	81.2	1386	22	AAH26362	DNA encoding human

## ALIGNMENTS

RESULT 1	AAAD15227	standard; DNA; 1386 BP.
AAAD15227	standard; DNA; 1386 BP.	
AAAD15227	standard; DNA; 1386 BP.	
01-NOV-2001	(First entry)	
Human protein C derivative encoding DNA #3.		
Human protein C derivative: anticoagulation activity; thrombosis;		
serpin inactivator; acute coronary syndrome; myocardial infarction;		
vascular occlusive disorder; hypercoagulable state; angina; sepsis;		
disseminated intravascular coagulation; DIC; burn; transplantation;		
sickle cell disease; viral haemorrhagic fever; protein C deficiency;		
haemolytic uremic syndrome; acute arterial thrombotic occlusion;		
thromboembolism; prothrombotic disorder; gene therapy; thalassemia; ds.		

```

OS Homo sapiens.
XX Key Location/Qualifiers
FH CDS 127..1386
FT /tag= a
FT /product= "Human protein C derivative"
FT /transl_except= (pos:154..156, aa:His)
XX
XX WO200159084-A1.
XX
XX 16-AUG-2001.
XX
XX 02-FEB-2001; 2001WO-US01221.
XX
XX 11-FEB-2000; 2000US-0181948.
XX
XX 14-MAR-2000; 2000US-0189199.
XX
XX (ELIL ) LILLY & CO ELI.
XX
XX Gerlitz BE, Grinnell BW, Jones BE;
XX
XX WPI; 2001-514662/56.
XX
XX P-PSDB; AAE08629.
XX
XX Protein C derivative for treating acute coronary syndromes, vascular
XX occlusive disorders, thrombotic disorders and sepsis, comprises
XX substitutions at specified amino acid positions
XX
XX
XX Disclosure: Page 54-55; 59pp; English.
XX
XX The invention relates to human protein C derivatives and nucleic acid
XX molecules encoding such derivatives. These derivatives have increased
XX anticoagulation activity, resistance to serpin inactivation and
XX increased sensitivity to thrombin activation compared to wild type
XX protein C, and retains the biological activity of the wild type human
XX protein C. Protein C derivatives are useful in the manufacture of a
XX medicament for the treatment of acute coronary syndromes e.g. myocardial
XX infarction and unstable angina; and disease states predisposing to
XX thrombosis; vascular occlusive disorders and hypercoagulable states e.g.
XX disseminated intravascular coagulation (DIC), burns, transplantations,
XX CC thalassemia, sickle cell disease, viral haemorrhagic fever and
XX CC haemolytic uremic syndrome; sepsis in combination with bacterial
XX permeability increasing protein; thrombotic disorders in combination
XX with an anti-platelet agent; protein C deficiency; acute arterial
XX CC thrombotic occlusion, thromboembolism or stenosis in coronary, cerebral
XX or peripheral arteries or in vascular grafts in combination with a
XX CC thrombolytic agent. Nucleic acid molecules of the invention are useful
XX for treating humans with genetically predisposed prothrombotic disorders
XX CC by gene therapy. The present sequence is human protein C derivative
XX encoding DNA.
XX
XX SQ Sequence 1386 BP; 286 A; 416 C; 443 G; 241 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 1,36e-23 Length: 1386
XX Score: 179.00 Matches: 35
XX Percent Similarity: 79.55% Conservative: 0
XX Best Local Similarity: 79.55% Mismatches: 9
XX Query Match: 90.86% Indels: 0
XX DB: 22 Gaps: 0
XX
XX SEQ1-4EDITS (1-44) x AAD15227 (1-1386)
XX
XX 1 AlaAsnSerPheLeu*****LeuArgInGlySerLeu***Arg**CysIle***** 20
XX |||||
XX Db 127 GCAACTCTCTCTGGAGGAGCTCCGTCAGAGGAGCCTGGAGCGGAGTGCATAGAGGAG 186
XX |||||
XX Qy 21 IleCysAspPhe*****AlaIys***IlePheGluAspValAspSphTrhLeuAlaPhe 40
XX |||||
XX Db 187 AFTCTGTGACTTCGAGAGAGCCAGAAATTTTCGAGAGATGTGATGACACACTGGCCTTC 246
XX |||||
XX Qy 41 TrpSerIysHis 44
XX |||||

```

```

Db 247 TGTGTCAAGCAC 258
XX
XX RESULT 2
XX AAD15228
XX ID AAD15228 standard; DNA: 1386 BP.
XX
XX AAD15228;
XX
XX 01-NOV-2001 (first entry)
XX
XX Human protein C derivative encoding DNA #4.
XX
XX DE
XX
XX Human; protein C derivative; anticoagulation activity; thrombosis;
XX KM serpin inactivation; acute coronary syndrome; myocardial infarction;
XX KM vascular occlusive disorder; hypercoagulable state; angina; sepsis;
XX KM disseminated intravascular coagulation; DIC; burn; transplantation;
XX KM sickle cell disease; viral haemorrhagic fever; protein C deficiency;
XX KM haemolytic uremic syndrome; acute arterial thrombotic occlusion;
XX KM thromboembolism; prothrombotic disorder; gene therapy; thalassemia; ds.
XX
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT CDS 127..1386
XX FT /tag= a
XX FT /product= "Human protein C derivative"
XX
XX
XX WO200159084-A1.
XX
XX 16-AUG-2001.
XX
XX 02-FEB-2001; 2001WO-US01221.
XX
XX 11-FEB-2000; 2000US-0181948.
XX
XX 14-MAR-2000; 2000US-0189199.
XX
XX (ELIL ) LILLY & CO ELI.
XX
XX Gerlitz BE, Grinnell BW, Jones BE;
XX
XX WPI; 2001-514662/56.
XX
XX P-PSDB; AAE08630.
XX
XX Protein C derivative for treating acute coronary syndromes, vascular
XX occlusive disorders, thrombotic disorders and sepsis, comprises
XX substitutions at specified amino acid positions
XX
XX
XX Disclosure: Page 55-56; 59pp; English.
XX
XX The invention relates to human protein C derivatives and nucleic acid
XX molecules encoding such derivatives. These derivatives have increased
XX anticoagulation activity, resistance to serpin inactivation and
XX increased sensitivity to thrombin activation compared to wild type
XX protein C, and retains the biological activity of the wild type human
XX protein C. Protein C derivatives are useful in the manufacture of a
XX medicament for the treatment of acute coronary syndromes e.g. myocardial
XX infarction and unstable angina; and disease states predisposing to
XX thrombosis; vascular occlusive disorders and hypercoagulable states e.g.
XX disseminated intravascular coagulation (DIC), burns, transplantations,
XX CC thalassemia, sickle cell disease, viral haemorrhagic fever and
XX CC haemolytic uremic syndrome; sepsis in combination with bacterial
XX permeability increasing protein; thrombotic disorders in combination
XX with an anti-platelet agent; protein C deficiency; acute arterial
XX CC thrombotic occlusion, thromboembolism or stenosis in coronary, cerebral
XX or peripheral arteries or in vascular grafts in combination with a
XX CC thrombolytic agent. Nucleic acid molecules of the invention are useful
XX for treating humans with genetically predisposed prothrombotic disorders
XX CC by gene therapy. The present sequence is human protein C derivative
XX encoding DNA.
XX
XX SQ Sequence 1386 BP; 286 A; 416 C; 443 G; 241 T; 0 other;
XX
XX
XX Alignment Scores:

```



Pred. No.: 1,36e-23 Length: 1386  
 Score: 179.00 Matches: 35  
 Percent Similarity: 79.55% Conservative: 0  
 Best Local Similarity: 79.55% Mismatches: 9  
 Query Match: 90.86% Indels: 0  
 DB: 22 Gaps: 0

SEQ1-4EDITS (1-44) x AAH26365 (1-1386)  
 Oy 1 AlaAsnSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 Db 127 GCCAAGCTCTTCCTGAGAGAGCTCCGTCAAGGAGCCTGGAGCGGAGTGCATAGAGAG 186  
 Oy 21 IleCysAspPhe\*\*\*\*\*AlaIlys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
 Db 187 ATCTGTGACTTCGAGAGAGCCAAAGAAATTTTCGAAGATGTGATGACACACTGGCCTTC 246  
 Oy 41 TrpSerLysHis 44  
 Db 247 TGGTCCAAAGCAC 258

RESULT 3  
 AAH26365  
 ID AAH26365 standard; cDNA: 1386 BP.  
 AC AAH26365:  
 DT 15-OCT-2001 (first entry)  
 XX  
 DE DNA encoding human protein C derivative.  
 XX  
 KW Protein C; human; coronary syndrome; thrombosis; angina;  
 KW myocardial infarction; vascular occlusive disorder;  
 KW hypercoagulatism; sepsis; protein C deficiency; occlusion;  
 KW thrombolytic; cardiant; antiangiinal; anticoagulant; gene therapy;  
 KW mutant; ss.  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT sig\_peptide 1..126  
 FT mat\_peptide 127..1383  
 FT /\*tag= b  
 FT /\*note= "encodes AAB82677"  
 FT  
 PN WO200157193-A2.  
 PD 09-AUG-2001.  
 XX  
 PF 19-JAN-2001; 2001WO-US00020.  
 XX  
 PR 02-FEB-2000; 2000US-0179801.  
 PR 14-MAR-2000; 2000US-0181917.  
 XX  
 PA (ELIL ) LILLY & CO ELI.  
 XX  
 PI Gerlitz BE, Jones BE;  
 XX  
 DR WPI: 2001-496919/54.  
 DR P-PSDB: AAB82677.  
 XX  
 PT Novel human protein C derivative for treating, e.g., myocardial  
 PT infarction, unstable angina, sepsis, thrombotic disorders, acute  
 PT arterial thrombotic occlusion, and thromboembolism .  
 XX  
 PS Disclosure: Page 60-61; 63pp; English.  
 CC The present sequence is that of DNA encoding a claimed human protein  
 CC C derivative (see AAB82677), in which His at position 10 of the  
 CC mature wild-type protein C sequence is substituted with Gln, Ser at

CC position 11 with Gly, Gln at position 32 with Glu, Asn at position  
 CC 33 with Asp and Leu at position 194 with Ser. The invention relates  
 CC to protein C derivatives having at least 2 amino acid substitutions,  
 CC and to recombinant DNA molecules encoding such derivatives. These  
 CC derivatives have increased anticoagulant activity and resistance to  
 CC inactivation by serpins compared with wild-type human protein C but  
 CC retain the biological activity of the wild-type protein. Recombinant  
 CC DNA molecules encoding preferred protein C derivatives are given in  
 CC AAH26363-66. Also claimed are a vector comprising the recombinant  
 CC DNA, transformed host cells and a method of producing the human  
 CC protein C derivative. The protein C derivatives are useful for  
 CC treating coronary syndromes and disease states predisposing to  
 CC thrombosis (e.g. myocardial infarction and unstable angina),  
 CC vascular occlusive disorders and hypercoagulable states, sepsis (in  
 CC combination with bactericidal permeability increasing protein or  
 CC with tissue factor pathway inhibitor), thrombotic disorders (in  
 CC combination with an anti-platelet agent or by local delivery through  
 CC an intracoronary catheter), protein C deficiency, acute arterial  
 CC thrombotic occlusion, thromboembolism, or stenosis in coronary,  
 CC cerebral or peripheral arteries or in vascular grafts. Human  
 CC patients with genetically predisposed prothrombotic disorders may  
 CC be treated by gene therapy (all claimed).  
 XX

SQ Sequence 1386 BP; 286 A; 416 C; 443 G; 241 T; 0 other:  
 Alignment Scores:  
 Pred. No.: 1,36e-23 Length: 1386  
 Score: 179.00 Matches: 35  
 Percent Similarity: 79.55% Conservative: 0  
 Best Local Similarity: 79.55% Mismatches: 9  
 Query Match: 90.86% Indels: 0  
 DB: 22 Gaps: 0

SEQ1-4EDITS (1-44) x AAH26365 (1-1386)  
 Oy 1 AlaAsnSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 Db 127 GCCAAGCTCTTCCTGAGAGAGCTCCGTCAAGGAGCCTGGAGCGGAGTGCATAGAGAG 186  
 Oy 21 IleCysAspPhe\*\*\*\*\*AlaIlys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
 Db 187 ATCTGTGACTTCGAGAGAGCCAAAGAAATTTTCGAAGATGTGATGACACACTGGCCTTC 246  
 Oy 41 TrpSerLysHis 44  
 Db 247 TGGTCCAAAGCAC 258

RESULT 4  
 AAH26366  
 ID AAH26366 standard; cDNA: 1386 BP.  
 AC AAH26366:  
 DT 15-OCT-2001 (first entry)  
 XX  
 DE DNA encoding human protein C derivative.  
 XX  
 KW Protein C; human; coronary syndrome; thrombosis; angina;  
 KW myocardial infarction; vascular occlusive disorder;  
 KW hypercoagulatism; sepsis; protein C deficiency; occlusion;  
 KW thrombolytic; cardiant; antiangiinal; anticoagulant; gene therapy;  
 KW mutant; ss.  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT sig\_peptide 1..126  
 FT mat\_peptide 127..1383  
 FT /\*tag= a  
 FT /\*tag= b  
 FT /\*note= "encodes AAB82678"  
 FT

XX WO200157193-A2.  
 XX 09-AUG-2001.  
 PD  
 XX  
 PF 19-JAN-2001; 2001WO-US00020.  
 XX  
 PR 02-FEB-2000; 2000US-0179801.  
 PR 14-MAR-2000; 2000US-0189197.  
 XX  
 PA (ELIL ) LILLY & CO ELI.  
 XX Gerlitz BE, Jones BE.  
 PI  
 DR WPI: 2001-496919/54.  
 XX P-PSDB; AAB82678.  
 XX  
 PT Novel human protein C derivative for treating, e.g., myocardial  
 PT infarction, unstable angina, sepsis, thrombotic disorders, acute  
 PT arterial thrombotic occlusion, and thromboembolism -  
 XX  
 PS Disclosure: Page 61-62; 63pp; English.  
 XX  
 CC The present sequence is that of DNA encoding a claimed human protein  
 CC C derivative (see AAB82678), in which His at position 10 of wild-type  
 CC protein C is substituted with Gln, Ser at position 11 with Gly, Gln  
 CC at position 32 with Glu, Asn at position 33 with Asp, Leu at position  
 CC 194 with Ser, and Thr at position 254 with Ser. The invention relates  
 CC to protein C derivatives having at least 2 amino acid substitutions,  
 CC and to recombinant DNA molecules encoding such derivatives. These  
 CC derivatives have increased anticoagulant activity and resistance to  
 CC inactivation by serpins compared with wild-type human protein C but  
 CC retain the biological activity of the wild-type protein. Recombinant  
 CC DNA molecules encoding preferred protein C derivatives are given in  
 CC AAB26363-66. Also claimed are a vector comprising the recombinant  
 CC DNA, transformed host cells and a method of producing the human  
 CC protein C derivative. The protein C derivatives are useful for  
 CC treating coronary syndromes and disease states predisposing to  
 CC thrombosis (e.g. myocardial infarction and unstable angina),  
 CC vascular occlusive disorders and hypercoagulable states, sepsis (in  
 CC combination with bactericidal permeability increasing protein or  
 CC with tissue factor pathway inhibitor), thrombotic disorders (in  
 CC combination with an anti-platelet agent or by local delivery through  
 CC an intracoronary catheter), protein C deficiency, acute arterial  
 CC thrombotic occlusion, thromboembolism, or stenosis in coronary,  
 CC cerebral or peripheral arteries or in vascular grafts. Human  
 CC patients with genetically predisposed prothrombotic disorders may  
 CC be treated by gene therapy (all claimed).  
 CC  
 XX  
 SQ Sequence 1386 BP; 286 A; 415 C; 444 G; 241 T; 0 other;

Alignment Scores:  
 Pred. No.: 1,366-23 Length: 1386  
 Score: 179.00 Matches: 35  
 Percent Similarity: 79.55% Conservative: 0  
 Best Local Similarity: 79.55% Mismatches: 9  
 Query Match: 90.86% Indels: 0  
 DB: 22 Gaps: 0

SEQ1-4EDITS (1-44) x AAB26366 (1-1386)

OY 1 AjaasSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 DB 127 GCCAACTCCTCTCGAGAGAGCTCGTCNAAGGAGCGCTGAGCGGAGTCATAGAGAG 186  
 OY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspPhrIleuAlaPhe 40  
 DB 187 ATCTGTGACTGTGAGAGGCGCAAGAAATTTTCAAGATGTGATGACACACTGGCTTC 246  
 OY 41 TTPSerLySHis 44  
 DB 247 TGGTCCAAAGCAC 258

RESULT 5  
 AAD15225  
 ID AAD15225 standard; DNA; 1386 BP.  
 XX  
 XX AAD15225;  
 AC  
 XX  
 DT 01-NOV-2001 (first entry)  
 XX  
 DE Human protein C derivative encoding DNA #1.  
 XX  
 KW Human; protein C derivative; anticoagulation activity; thrombosis;  
 KW serpin inactivation; acute coronary syndrome; myocardial infarction;  
 KW vascular occlusive disorder; hypercoagulable state; angina; sepsis;  
 KW disseminated intravascular coagulation; DIC; burn; transplantation;  
 KW sickle cell disease; viral haemorrhagic fever; protein C deficiency;  
 KW haemolytic uremic syndrome; acute arterial thrombotic occlusion;  
 KW thromboembolism; prothrombotic disorder; gene therapy; thalassemia; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 127..1386  
 FT /tag= a  
 FT /product= "Human protein C derivative"  
 XX  
 PN WO200159084-A1.  
 PD 16-AUG-2001.  
 XX  
 PF 02-FEB-2001; 2001WO-US01221.  
 XX  
 PR 11-FEB-2000; 2000US-0181948.  
 PR 14-MAR-2000; 2000US-0189199.  
 XX  
 PA (ELIL ) LILLY & CO ELI.  
 PI Gerlitz BE, Grinnell BW, Jones BE;  
 XX  
 DR WPI: 2001-514662/56.  
 DR P-PSDB; AAE08627.  
 XX  
 PT Protein C derivative for treating acute coronary syndromes, vascular  
 PT occlusive disorders, thrombotic disorders and sepsis, comprises  
 PT substitutions at specified amino acid positions -  
 XX  
 XX Disclosure: Page 53; 59pp; English.  
 CC The invention relates to human protein C derivatives and nucleic acid  
 CC molecules encoding such derivatives. These derivatives have increased  
 CC anticoagulation activity, resistance to serpin inactivation and  
 CC increased sensitivity to thrombin activation compared to wild type  
 CC protein C, and retains the biological activity of the wild type human  
 CC protein C. Protein C derivatives are useful in the manufacture of a  
 CC medicament for the treatment of acute coronary syndromes e.g. myocardial  
 CC infarction and unstable angina; and disease states predisposing to  
 CC thrombosis; vascular occlusive disorders and hypercoagulable states e.g.  
 CC disseminated intravascular coagulation (DIC), burns, transplantations,  
 CC thalassemia, sickle cell disease, viral haemorrhagic fever and  
 CC haemolytic uremic syndrome; sepsis in combination with bacterial  
 CC permeability increasing protein; thrombotic disorders in combination  
 CC with an anti-platelet agent; protein C deficiency; acute arterial  
 CC thrombotic occlusion, thromboembolism or stenosis in coronary, cerebral  
 CC or peripheral arteries or in vascular grafts in combination with a  
 CC thrombolytic agent. Nucleic acid molecules of the invention are useful  
 CC for treating humans with genetically predisposed prothrombotic disorders  
 CC by gene therapy. The present sequence is human protein C derivative  
 CC encoding DNA.  
 XX  
 SQ Sequence 1386 BP; 285 A; 418 C; 442 G; 241 T; 0 other;

Alignment Scores:  
 Pred. No.: 1,196-22 Length: 1386  
 Score: 174.00 Matches: 34

```
Percent Similarity: 77.27%      Conservative: 0
Best Local Similarity: 77.27%    Mismatches: 10
Query Match: 88.32%             Indels: 0
DB: 22                           Gaps: 0

SEQ1-4EDITS (1-44) x AAD15226 (1-1386)

QY 1 AlaAsnSerPheLeu*****LeuArgGlnGlySerLeu***Arg***CysIle***** 20
DB 127 GCCAACTCCTTCCTGAGAGAGCTCCGTCAAGGAGCTGAGAGCGGAGGAGCTATAGAGAG 186
QY 21 IleCysAspPhe*****AlaIlys***IlePheGluAspValAspAspThrLeuAlaPhe 40
DB 187 ATCTGTGACTTCGAGAGGCGCAAGCAATTTTCAAGATGTGTGATGACACACTGGCCTTC 246
QY 41 TrpSerLysHis 44
DB 247 TGGTCCAAGCAC 258

RESULT 6
AAD15226
ID AAD15226 standard; DNA: 1386 BP.
XX AC AAD15226:
XX AD15226:
XX 01-NOV-2001 (first entry)
XX DE Human protein C derivative encoding DNA #2.
XX KW Human: protein C derivative; anticoagulation activity; thrombosis;
XX KW serpin inactivation; acute coronary syndrome; myocardial infarction;
XX KW vascular occlusive disorder; hypercoagulable state; angina; sepsis;
XX KW disseminated intravascular coagulation; DIC; burn; transplantation;
XX KW sickle cell disease; viral haemorrhagic fever; protein C deficiency;
XX KW haemolytic uremic syndrome; acute arterial thrombotic occlusion;
XX KW thromboembolism; prothrombotic disorder; gene therapy; thalassaemia; ds.
XX OS Homo sapiens.
XX FH Key location/Qualifiers
XX FT CDS 127..1386
XX FT /*tag= a
XX FT /product= "Human protein C derivative"
XX PN MO200159084-A1.
XX PD 16-AUG-2001.
XX PE 02-FEB-2001: 2001WO-US01221.
XX PR 11-FEB-2000: 2000US-0181948.
XX PR 14-MAR-2000: 2000US-0189199.
XX PA (ELIL ) LILLY & CO ELI.
XX PI Gerlitz BE, Grinnell BW, Jones BE;
XX WP: 2001-514662/56.
XX DR P-PSDB: AAE08628.
XX PT Protein C derivative for treating acute coronary syndromes, vascular
XX PT occlusive disorders, thrombotic disorders and sepsis, comprises
XX PT substitutions at specified amino acid positions
XX PS Disclosure: Page 53-54; 59pp; English.
XX CC The invention relates to human protein C derivatives and nucleic acid
XX CC molecules encoding such derivatives. These derivatives have increased
XX CC anticoagulation activity, resistance to serpin inactivation and
XX CC increased sensitivity to thrombin activation compared to wild type
XX CC protein C, and retains the biological activity of the wild type human
XX CC protein C. Protein C derivatives are useful in the manufacture of a
XX CC medicament for the treatment of acute coronary syndromes e.g. myocardial
```

```
CC infection and unstable angina; and disease states predisposing to
CC thrombosis; vascular occlusive disorders and hypercoagulable states e.g.
CC disseminated intravascular coagulation (DIC), burns, transplantations,
CC thalassaemia, sickle cell disease, viral haemorrhagic fever and
CC haemolytic uremic syndrome; sepsis in combination with bacterial
CC permeability increasing protein; thrombotic disorders in combination
CC with an anti-platelet agent; protein C deficiency; acute arterial
CC thrombotic occlusion, thromboembolism or stenosis in coronary, cerebral
CC or peripheral arteries or in vascular grafts in combination with a
CC thrombolytic agent. Nucleic acid molecules of the invention are useful
CC for treating humans with genetically predisposed prothrombotic disorders
CC by gene therapy. The present sequence is human protein C derivative
CC encoding DNA.
XX.

SQ Sequence 1386 BP: 285 A; 417 C; 443 G; 241 T; 0 other:

Alignment Scores:
Pred. No.: 1,19e-22      Length: 1386
Score: 174.00           Matches: 34
Percent Similarity: 77.27%      Conservative: 0
Best Local Similarity: 77.27%    Mismatches: 10
Query Match: 88.32%           Indels: 0
DB: 22                       Gaps: 0

SEQ1-4EDITS (1-44) x AAD15226 (1-1386)

QY 1 AlaAsnSerPheLeu*****LeuArgGlnGlySerLeu***Arg***CysIle***** 20
DB 127 GCCAACTCCTTCCTGAGAGAGCTCCGTCAAGGAGCTGAGAGCGGAGGAGCTATAGAGAG 186
QY 21 IleCysAspPhe*****AlaIlys***IlePheGluAspValAspAspThrLeuAlaPhe 40
DB 187 ATCTGTGACTTCGAGAGGCGCAAGCAATTTTCAAGATGTGTGATGACACACTGGCCTTC 246
QY 41 TrpSerLysHis 44
DB 247 TGGTCCAAGCAC 258

RESULT 7
AAH26363
ID AAH26363 standard; cDNA: 1386 BP.
XX AC AAH26363:
XX AD26363:
XX 15-OCT-2001 (first entry)
XX DE DNA encoding human protein C derivative.
XX KW Protein C; human; coronary syndrome; thrombosis; angina;
XX KW myocardial infarction; vascular occlusive disorder;
XX KW hypercoagulation; sepsis; protein C deficiency; occlusion;
XX KW thromboembolism; stenosis; antibacterial; immunosuppressive;
XX KW thrombolytic; cardiant; antianginal; anticoagulant; gene therapy;
XX KW mutant; ss.
XX OS Homo sapiens.
XX FH Key location/Qualifiers
XX FT sig_peptide 1..126
XX FT /*tag= a
XX FT mat_peptide 127..1383
XX FT /*tag= b
XX FT /note= "encodes AAB82675"
XX PN WO200157193-A2.
XX PD 09-AUG-2001.
XX PE 19-JAN-2001: 2001WO-US00020.
XX PR 02-FEB-2000: 2000US-0179801.
XX PR 14-MAR-2000: 2000US-0189197.
```

```

XX (ELIL ) LILLY & CO ELI.
XX
XX Gerlitz BE, Jones BE;
XX
XX WPI: 2001-496919/54.
XX P-PSDB: AAB82675.
XX
XX Novel human protein C derivative for treating, e.g., myocardial
XX infection, unstable angina, sepsis, thrombotic disorders, acute
XX arterial thrombotic occlusion, and thromboembolism -
XX
XX Disclosure: Page 58-59; 63pp; English.
XX
XX The present sequence is that of DNA encoding a claimed human
XX protein C derivative (see AAB82675), in which Ser at position 11 of
XX the mature wild-type protein C sequence is substituted with Gly,
XX Gln at position 32 with Glu, Asn at position 33 with Asp, and Leu
XX at position 194 with Ser. The invention relates to protein C
XX derivatives having at least 2 amino acid substitutions, and to
XX recombinant DNA molecules encoding such derivatives. These
XX derivatives have increased anticoagulant activity and resistance to
XX inactivation by serpins compared with wild-type human protein C but
XX retain the biological activity of the wild-type protein. Recombinant
XX DNA molecules encoding preferred protein C derivatives are given in
XX CC AAB26363-66. Also claimed are a vector comprising the recombinant
XX CC DNA, transformed host cells and a method of producing the human
XX CC protein C derivative. The protein C derivatives are useful for
XX CC treating coronary syndromes and disease states predisposing to
XX CC thrombosis (e.g. myocardial infarction and unstable angina),
XX CC vascular occlusive disorders and hypercoagulable states, sepsis (in
XX CC combination with bactericidal permeability increasing protein or
XX CC with tissue factor pathway inhibitor), thrombotic disorders (in
XX CC combination with an anti-platelet agent or by local delivery through
XX CC an intracoronary catheter), protein C deficiency, acute arterial
XX CC cerebral or peripheral arteries or in vascular grafts. Human
XX CC patients with genetically predisposed prothrombotic disorders may
XX CC be treated by gene therapy (all claimed).
XX
XX SQ Sequence 1386 BP; 285 A; 417 C; 443 G; 241 T; 0 other:
XX
XX Alignment Scores:
XX Pred. No.: 1.19e-22 Length: 1386
XX Score: 174.00 Matches: 34
XX Percent Similarity: 77.27% Conservative: 0
XX Best Local Similarity: 77.27% Mismatches: 10
XX Query Match: 88.32% Indels: 0
XX Gaps: 0
XX
XX SE01-4EDITS (1-44) x AAB26363 (1-1386)
XX
XX 1 AlaAsnSerPheLeu*****LeuArgGlnGlySerLeu**Arg**CysIle***** 20
XX ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 127 GCCAACTCTTCTCTGAGAGAGCTCCGTCACGGAGCCTTGAGCGGAGTGCATAGAGGAG 186
XX
XX 21 IleCysAspPhe*****AlaIlys***IlePheGluAspValaIAspAspThrLeuAlaPhe 40
XX ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX Db 187 AACTGTGAGCTTGAGAGGCGCAAGAAATTTCGAAAGATGTGATGACACACTGGCCTTC 246
XX
XX 41 TTPSerLyHis 44
XX ||||||| |||||
XX Db 247 TGGTCCAAAGCAC 258
XX
XX RESULT 8
XX ID AAB26364 standard: cDNA; 1386 BP.
XX
XX AAH26364;
XX
XX AC AAH26364;
XX
XX DT 15-OCT-2001 (first entry)
XX
XX XX DNA encoding human protein C derivative.
XX
XX DE

```

```

XX Protein C; human; coronary syndrome; thrombosis; angina;
XX myocardial infarction; vascular occlusive disorder;
XX hypercoagulation; sepsis; protein C deficiency; occlusion;
XX thromboembolism; stenosis; antibacterial; immunosuppressive;
XX thrombolytic; cardiac; antiangiinal; anticoagulant; gene therapy;
XX mutant; ss.
XX
XX Homo sapiens.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX sig_peptide 1..126
XX FT /*tag= a
XX FT 127..1383
XX FT /*tag= b
XX FT /note= "encodes AAB82676"
XX
XX WO200157193-A2.
XX
XX 09-AUG-2001.
XX
XX 19-JAN-2001; 2001WO-US00020.
XX
XX 02-FEB-2000; 2000US-0179801.
XX
XX 14-MAR-2000; 2000US-0189197.
XX
XX (ELIL ) LILLY & CO ELI.
XX
XX Gerlitz BE, Jones BE;
XX
XX WPI: 2001-496919/54.
XX P-PSDB: AAB82676.
XX
XX Novel human protein C derivative for treating, e.g., myocardial
XX infection, unstable angina, sepsis, thrombotic disorders, acute
XX arterial thrombotic occlusion, and thromboembolism -
XX
XX Disclosure: Page 59-60; 63pp; English.
XX
XX The present sequence is that of DNA encoding a claimed human
XX protein C derivative (see AAB82676), in which Ser at position 11 of
XX the mature wild-type protein C sequence is substituted with Gly, Gln
XX at position 32 with Glu, Asn at position 33 with Asp, Leu at position
XX 194 with Ser, and Thr at position 254 with Ser. The invention relates
XX to protein C derivatives having at least 2 amino acid substitutions,
XX and to recombinant DNA molecules encoding such derivatives. These
XX derivatives have increased anticoagulant activity and resistance to
XX inactivation by serpins compared with wild-type human protein C but
XX retain the biological activity of the wild-type protein. Recombinant
XX DNA molecules encoding preferred protein C derivatives are given in
XX CC AAB26363-66. Also claimed are a vector comprising the recombinant
XX CC DNA, transformed host cells and a method of producing the human
XX CC protein C derivative. The protein C derivatives are useful for
XX CC treating coronary syndromes and disease states predisposing to
XX CC thrombosis (e.g. myocardial infarction and unstable angina),
XX CC vascular occlusive disorders and hypercoagulable states, sepsis (in
XX CC combination with bactericidal permeability increasing protein or
XX CC with tissue factor pathway inhibitor), thrombotic disorders (in
XX CC combination with an anti-platelet agent or by local delivery through
XX CC an intracoronary catheter), protein C deficiency, acute arterial
XX CC thrombotic occlusion, thromboembolism, or stenosis in coronary,
XX CC cerebral or peripheral arteries or in vascular grafts. Human
XX CC patients with genetically predisposed prothrombotic disorders may
XX CC be treated by gene therapy (all claimed).
XX
XX SQ Sequence 1386 BP; 285 A; 416 C; 444 G; 241 T; 0 other:
XX
XX Alignment Scores:
XX Pred. No.: 1.19e-22 Length: 1386
XX Score: 174.00 Matches: 34
XX Percent Similarity: 77.27% Conservative: 0
XX Best Local Similarity: 77.27% Mismatches: 10
XX

```

Query Match: 88.32% Indels: 0  
 DB: 22 Gaps: 0  
 SEQ1-4EDITS (1-44) x AAH26364 (1-1386)

OY 1 AlaasSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 DB 127 GCCAACTCTCTCTCGAGAGAGCTCCGTCACAGCGAGCTCGAGCGGAGTGCATAGAGGAG 186  
 OY 21 IleCysAspPhe\*\*\*\*\*AlaLys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
 DB 187 ATCTGTGACTTGTGAGAGGCGCAAGAAATTTTCGAAGATGTGATGACACACTGGCCCTTC 246

OY 41 TrpSerLysHis 44  
 DB 247 TGGTCCAAAGCAGC 258

RESULT 9  
 AAZ46750  
 ID AAZ46750 standard: DNA: 1245 BP.  
 XX  
 AC AAZ46750:  
 XX  
 DT 27-MAR-2000 (first entry)  
 DE Truncated human protein C polypeptide encoding DNA.  
 XX  
 KW Protein C; truncated; thrombotic disorder; vascular disorder; stroke;  
 KW hypercoagulable state; myocardial infarction; unstable angina; sepsis;  
 KW adult respiratory distress syndrome; sickle cell anemia; human; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9963070-A1.  
 PD 09-DEC-1999.  
 XX  
 PF 01-JUN-1999; 99WO-US11969.  
 XX  
 PR 01-JUN-1998; 98US-0087585.  
 XX  
 PA (ELI ) LILLY & CO ELI.  
 PI Huang L, Riggin RM;  
 DR WPI: 2000-086975/07.  
 DR P-PSDB: AAY56803.  
 XX  
 PT Novel polypeptide useful for treating thrombotic and vascular diseases  
 PT and hypercoagulation, e.g. stroke  
 XX  
 PS Claim 4: Page 21: 23pp; English.  
 XX  
 CC This DNA encodes a human protein C polypeptide having a light chain and  
 CC a truncated heavy chain. The protein can be produced by standard  
 CC recombinant methodologies. The truncated protein C is used to treat a  
 CC wide range of thrombotic or vascular disorders or hypercoagulable states,  
 CC e.g. stroke; myocardial infarction; unstable angina; sepsis; adult  
 CC respiratory distress syndrome; sickle cell anemia etc. The truncated  
 CC protein C retains the activity of full-length protein C but does not  
 CC undergo C-terminal cleavage, of the heavy chain, during activation.  
 XX  
 SO Sequence 1245 BP: 262 A; 371 C; 402 G; 210 T; 0 other;

Alignment Scores:  
 Pred. No.: 4.46e-20 Length: 1245  
 Score: 160.00 Matches: 31  
 Percent Similarity: 75.00% Conservative: 2  
 Best Local Similarity: 70.45% Mismatches: 11  
 Query Match: 81.22% Indels: 0  
 DB: 21 Gaps: 0

SEQ1-4EDITS (1-44) x AAZ46750 (1-1245)

OY 1 AlaasSerPheLeu\*\*\*\*\*LeuAArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 DB 1 GCCAACTCTCTCTCGAGAGAGCTCCGTCACAGCAGCTCGAGCGGAGTGCATAGAGGAG 60  
 OY 21 IleCysAspPhe\*\*\*\*\*AlaLys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
 DB 61 ATCTGTGACTTGTGAGAGGCGCAAGAAATTTTCGAAGATGTGATGACACACTGGCCCTTC 120

OY 41 TrpSerLysHis 44  
 DB 121 TGGTCCAAAGCAGC 132

RESULT 10  
 ABR86039  
 ID ABR86039 standard: DNA: 1257 BP.  
 XX  
 AC ABR86039:  
 XX  
 DT 23-AUG-2002 (first entry)  
 DE Synthetic DNA encoding zymogen protein C.  
 XX  
 KW Human; Protein C; zymogen protein; ds; gene; N-glycosylation;  
 KW serum half-life; chromosome 2q13-q14; stroke; myocardial infarction;  
 KW after venous thrombosis; disseminated intravascular coagulation; DIC;  
 KW sepsis; septic shock; embolism; pulmonary embolism; burn; pregnancy;  
 KW bone marrow transplantation; major surgery; trauma; AIDS; coagulant;  
 KW adult respiratory distress syndrome; alpha-1 antitrypsin; APC;  
 KW activated protein C.  
 XX  
 OS Homo sapiens.  
 XX  
 PN Synthetic.  
 OS  
 FH Key Location/Qualifiers  
 FT 1..1257  
 FT CDS  
 FT /\*tag= a  
 FT /product= "Zymogen protein C"  
 FT /partial  
 FT /note= "No start or stop codon shown"

WO200232461-A2.  
 PN  
 PD 25-APR-2002.  
 XX  
 PF 15-OCT-2001; 2001WO-DK00679.  
 XX  
 PR 18-OCT-2000; 2000DK-0001560.  
 PR 18-OCT-2000; 2000US-242268P.  
 PR 21-JUN-2001; 2001DK-0000970.  
 PR 21-JUN-2001; 2001US-300154P.  
 XX  
 PA (MAXY-) MAXYGEN APS.  
 PA (MAXY-) MAXYGEN HOLDINGS LTD.  
 PI Andersen KV, Pedersen AH, Freskgaard PO;  
 DR WPI: 2002-489675/52.  
 DR P-PSDB: AAU86039.  
 XX  
 PT Novel conjugate useful for treating or preventing septic shock, stroke  
 PT and myocardial infarction, comprises non-polypeptide group covalently  
 PT attached to protein C polypeptide comprising an attachment group  
 XX  
 PS Disclosure: Page 78-79; 92pp; English.  
 XX  
 CC The invention relates to a conjugate (I) comprising at least one non-  
 CC polypeptide moiety (II) (e.g. an N-glycosyl group) covalently attached to  
 CC a protein C polypeptide comprising an amino acid sequence which differs  
 CC from that of a parent protein C polypeptide (III) in at least one  
 CC introduced and/or at least one removed amino acid residue comprising an  
 CC attachment group for the non-polypeptide group (e.g. an N-glycosylation  
 CC site). Also included are (i) a variant (IV) of (III) comprising a

CC substitution in a position (P) where (P) is an amino acid with at  
CC least 25% of its side group exposed to the surface, with the proviso that  
CC the substitution is not Thr245Ser/Ala/His/Lys/Arg/Asn/Asp/Glu/Gly/Gln,  
CC Tyr302Ser/Ala/Thr/His/Lys/Arg/Asn/Asp/Glu/Gly/Gln or Phe316Ser/Ala/Thr/  
CC His/Lys/Arg/Asn/Asp/Glu/Gly/Gln; (2) a nucleotide sequence (V) encoding  
CC (IV); (3) an expression vector (VI) comprising (V); (4) a host cell (VII)  
CC comprising (V) or (VI); (5) increasing (M2) the functional in vivo half-  
CC life or the serum half-life of a parent protein C polypeptide.  
CC The conjugates, variants and protein C proteins are useful as  
CC medicaments, and in the manufacture of medicaments for the treatment (and  
CC diagnosis/prevention) of stroke, myocardial infarction, after venous  
CC thrombosis, disseminated intravascular coagulation (DIC), sepsis, septic  
CC shock, emboli e.g. pulmonary emboli, transplantation such as bone marrow  
CC transplantation, burns, pregnancy, major surgery/trauma or adult  
CC respiratory distress syndrome (ARDS). The variant protein C has an  
CC increased resistance to activation by e.g. human plasma and alpha-1  
CC antitrypsin. The conjugates have an increased in vivo half-life,  
CC increased serum half-life, increased resistance to inhibitors, reduced  
CC renal clearance, reduced immunogenicity and/or increased bioavailability.  
CC The conjugate offers a number of advantages over the currently available  
CC APC products, including longer duration between injections.  
CC administration of less protein, and fewer side effects. Moreover, a  
CC reduced anticoagulant activity is beneficial to reduce the risk of  
CC bleeding while maintaining the antiinflammatory activity of APC  
CC (activated protein C) conjugates. This must be especially important when  
CC the conjugate has an extended plasma life. The gene for protein C is  
CC located on chromosome 2q13-q14. The present sequence encodes zymogen  
CC protein C.

SO Sequence 1257 BP; 264 A; 375 C; 406 G; 212 T; 0 other;

Alignment Scores:  
Pred. No.: 4.51e-20 Length: 1257  
Score: 160.00 Matches: 31  
Percent Similarity: 75.00% Conservative: 2  
Best Local Similarity: 70.45% Mismatches: 11  
Query Match: 81.22% Indels: 0  
DB: 24 Gaps: 0

SEQ1-4EDITS (1-44) x ABR86039 (1-1257)

OY 1 AlaAsnSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
DB 1 GCCAACTCCTCTCTCGAGAGAGCTCCGTCACAGCAGCGCTGGAGGGAGTGCATAGAGGAG 60  
OY 21 IleCysAspPhe\*\*\*\*\*AlaLys\*\*\*IlePheGluAspValAspAspHrLeuAlaPhe 40  
DB 61 ATCTGTGACTTTCGAGAGGCGCAAGAAATTTCCTCAAAATGTGATGATCACACACGCGCTTC 120  
OY 41 TTPSerLysHis 44  
DB 121 TGTGTCCAGCAC 132

RESULT 11

ID AAN60001 standard; DNA; 1260 BP.

XX AAN60001;

DT 25-JUL-1991 (first entry)

XX Sequence encoding polypeptide with human protein C activity.

XX Vascular disorder therapy; protein C deficiency; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..1260

PN EP191606-A.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PD 20-AUG-1986.

XX 06-FEB-1986; 86EP-0300823.

XX 08-FEB-1985; 85US-0699967.

XX (ELIL) ELI LILLY & CO.

PI Bang NU, Beckmann RJ, Jaskunas SR, Lai MHT, Little SP;

PI Long GL, Santerre RF;

DR WPI; 1986-220077/34.

DR P-PSDB; AAB60001.

PT Prod. of polypeptide having human protein C activity - is by

PT recombinant DNA procedures for prod. useful against vascular

PT disorders

PS Claim 1; Pages 96-97; 121pp; English.

CC The claimed sequence AAN60001 has "R1N-RM" attached to its 5' end

CC wherein: R= AAN60002 or AAN60003, and R1= AAN60004 or AAN60005; and M and

CC N= 0 or 1; provided that when M=0, N=0; and that when R= AAN60002, R1=

CC AAN60004; and that when R= AAN60003, R1= AAN60005.

XX Sequence 1260 BP; 265 A; 375 C; 407 G; 213 T; 0 other;

SO Sequence 1260 BP; 265 A; 375 C; 407 G; 213 T; 0 other;

Alignment Scores:  
Pred. No.: 4.52e-20 Length: 1260  
Score: 160.00 Matches: 31  
Percent Similarity: 75.00% Conservative: 2  
Best Local Similarity: 70.45% Mismatches: 11  
Query Match: 81.22% Indels: 0  
DB: 7 Gaps: 0

SEQ1-4EDITS (1-44) x AAN60001 (1-1260)

OY 1 AlaAsnSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
DB 1 GCCAACTCCTCTCTCGAGAGAGCTCCGTCACAGCAGCGCTGGAGGGAGTGCATAGAGGAG 60  
OY 21 IleCysAspPhe\*\*\*\*\*AlaLys\*\*\*IlePheGluAspValAspAspHrLeuAlaPhe 40  
DB 61 ATCTGTGACTTTCGAGAGGCGCAAGAAATTTCCTCAAAATGTGATGATCACACACGCGCTTC 120  
OY 41 TTPSerLysHis 44  
DB 121 TGTGTCCAGCAC 132

RESULT 12

ID AAD15223 standard; DNA; 1260 BP.

XX AAD15223;

DT 01-NOV-2001 (first entry)

XX Human mature wild type protein C encoding DNA.

XX Human; protein C derivative; anticoagulation activity; thrombosis;

XX serpin inactivation; acute coronary syndrome; myocardial infarction;

XX disseminated intravascular coagulation; DIC; burn; transplantation;

XX single cell disease; viral haemorrhagic fever; protein C deficiency;

XX haemolytic uremic syndrome; acute arterial thrombotic occlusion;

XX thromboembolism; prothrombotic disorder; gene therapy; thalassemia; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1..1260

FT /tag= a

FT /product= "Human mature wild type protein C"



Query Match: 81.22% Indels: 0  
 DB: 22 Gaps: 0  
 SEQ1-4EDITS (1-44) x AAH26361 (1-1260)  
 QY 1 AlaAsnSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 DB 1 GCCAACTCCTCTCGAGAGAGCCGTCACAGACGCTGGAGCGGAGTGCATGAGAGAG 60  
 QY 21 IleCysAspPhe\*\*\*\*\*AlaIlys\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
 DB 61 ATCTGTACTCTTCGAGAGAGCCCAAGAAATTTTCCAAATGTGATGACACACTGGCCTTC 120  
 QY 41 TrpSerLysHis 44  
 DB 121 TGGTCCAAGCAC 132  
 RESULT 14  
 AAC83311  
 ID AAC83311 standard; DNA: 1260 BP.  
 AC XX  
 AC AAC83311:  
 XX  
 DT 26-FEB-2001 (first entry)  
 XX  
 DE DNA encoding human protein C derivative 1.  
 XX  
 KM Protein C; human; vascular occlusive; burn; transplantation;  
 KM deep vein thrombosis; sickle cell; thalassemia; angina; stroke; ds.  
 KM thrombotic disorders; myocardial infarction;  
 OS Homo sapiens.  
 XX  
 PN WO200066754-A1.  
 XX  
 PD 09-NOV-2000.  
 XX  
 PF 13-APR-2000; 2000WO-0508722.  
 XX  
 PR 30-APR-1999; 99US-0131801.  
 XX  
 PA (ELIL ) LILLY & CO ELI.  
 XX  
 PI Gerilitz BE, Jones BE;  
 XX  
 DR WPI: 2001-007227/01.  
 DR P-PSDB; AAB36894.  
 PT Protein C derivatives, useful for treating vascular occlusive disorder,  
 PT hypercoagulable state, thrombotic disorder and disease states  
 PT predisposing thrombosis, comprises specific amino acid substitutions -  
 XX  
 PS Disclosure: Page 51-52; 57pp; English.  
 XX  
 CC The present invention relates to a human protein C derivative. The  
 CC protein is useful for treating vascular occlusive disorders,  
 CC hypercoagulable states such as sepsis, disseminated intravascular  
 CC coagulation, purpura fulminans, major trauma, major surgery, burns,  
 CC adult respiratory distress syndrome, transplantation, deep vein  
 CC thrombosis, heparin-induced thrombocytopenia, sickle cell disease,  
 CC thalassemia, viral hemorrhagic fever, thrombotic thrombocytopenic  
 CC purpura, and hemolytic uremic syndrome, and also useful for treating  
 CC thrombotic disorders and acute coronary syndromes such as myocardial  
 CC infarction, unstable angina, and stroke. Protein C derivatives with  
 CC amino acid substitutions result in increased resistance to  
 CC inactivation by serpins when compared to wild-type activated human  
 CC protein C. They also have longer half-lives in human blood and hence  
 CC require either less frequent administration and/or smaller dosage  
 CC than wild type human protein C for treating disorders.  
 XX  
 SO Sequence 1260 BP; 265 A; 375 C; 407 G; 213 T; 0 other;  
 Alignment Scores:

Pred. No.: 4,52e-20 Length: 1260  
 Score: 160.00 Matches: 31  
 Percent Similarity: 75.00% Conservative: 2  
 Best Local Similarity: 70.45% Mismatches: 11  
 Query Match: 81.22% Indels: 0  
 DB: 22 Gaps: 0  
 SEQ1-4EDITS (1-44) x AAC83311 (1-1260)  
 QY 1 AlaAsnSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 DB 1 GCCAACTCCTCTCGAGAGAGCCGTCACAGACGCTGGAGCGGAGTGCATGAGAGAG 60  
 QY 21 IleCysAspPhe\*\*\*\*\*AlaIlys\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
 DB 61 ATCTGTACTCTTCGAGAGAGCCCAAGAAATTTTCCAAATGTGATGACACACTGGCCTTC 120  
 QY 41 TrpSerLysHis 44  
 DB 121 TGGTCCAAGCAC 132  
 RESULT 15  
 ABR86038  
 ID ABR86038 standard; DNA: 1383 BP.  
 XX  
 AC ABR86038:  
 XX  
 DT 23-AUG-2002 (first entry)  
 XX  
 DE Synthetic DNA encoding protein C precursor protein.  
 XX  
 KM Human; Protein C; precursor protein; ds; gene; N-glycosylation;  
 KM serum half-life; chromosome 2q13-q14; stroke; myocardial infarction;  
 KM after venous thrombosis; disseminated intravascular coagulation; DIC;  
 KM sepsis; septic shock; embolism; pulmonary embolism; burn; pregnancy;  
 KM bone marrow transplantation; major surgery; trauma; ARDS; coagulant;  
 KM adult respiratory distress syndrome; alpha-1 antitrypsin; APC;  
 KM activated protein C.  
 OS Homo sapiens.  
 XX  
 OS Synthetic.  
 XX  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH 1..1383  
 FT CDS  
 FT FT  
 FT FT  
 FT FT  
 FT sig\_peptide  
 FT 1..126  
 FT /\*tag= b  
 FT mat\_peptide 127..1383  
 FT /\*tag= c  
 FT /product= Mature\_protein\_C  
 XX  
 XX WO200232461-A2.  
 XX  
 XX 25-APR-2002.  
 XX  
 PF 15-OCT-2001; 2001WO-DK00679.  
 XX  
 XX 18-OCT-2000; 2000DK-0001560.  
 PR 18-OCT-2000; 2000US-242268P.  
 PR 21-JUN-2001; 2001DK-0000970.  
 PR 21-JUN-2001; 2001US-300154P.  
 XX  
 XX (MAXY-) MAXYGEN APS.  
 PA (MAXY-) MAXYGEN HOLDINGS LTD.  
 XX  
 XX Andersen KV, Pedersen AH, Freskgaard PO;  
 PI  
 DR WPI: 2002-489875/52.  
 DR P-PSDB; AA099001.  
 XX



Novel conjugate useful for treating or preventing septic shock, stroke and myocardial infarction, comprises non-polypeptide group covalently attached to protein C polypeptide comprising an attachment group -

Example 4; Page 74-76; 92pp: English.

The invention relates to a conjugate (I) comprising at least one non-polypeptide moiety (II) (e.g. an N-glycosyl group) covalently attached to a protein C polypeptide comprising an amino acid sequence which differs from that of a parent protein C polypeptide (III) in at least one introduced and/or at least one removed amino acid residue comprising an attachment group for the non-polypeptide group (e.g. an N-glycosylation site). Also included are (1) a variant (IV) of (III) comprising a substitution in a position (P) where (P) is an amino acid with at least 25% of its side group exposed to the surface, with the proviso that the substitution is not Thr455Ser/Ala/His/Lys/Arg/Asn/Asp/Glu/Gly/Gln, Tyr302Ser/Ala/Thr/His/Lys/Arg/Asn/Asp/Glu/Gly/Gln or Phe316Ser/Ala/Thr/His/Lys/Arg/Asn/Asp/Glu/Gly/Gln; (2) a nucleotide sequence (V) encoding (IV); (3) an expression vector (VI) comprising (V); (4) a host cell (VII) comprising (V) or (VI); (5) increasing (M2) the functional in vivo half-life of the serum half-life of a parent protein C polypeptide. The conjugates, variants and protein C proteins are useful as medicaments, and in the manufacture of medicaments for the treatment (and diagnosis/prevention) of stroke, myocardial infarction, after venous thrombosis, disseminated intravascular coagulation (DIC), sepsis, septic shock, emboli e.g. pulmonary emboli, transplantation such as bone marrow transplantation, burns, pregnancy, major surgery/trauma or adult respiratory distress syndrome (ARDS). The variant protein C has an increased resistance to activation by e.g. human plasma and alpha-1 antitrypsin. The conjugates have an increased in vivo half-life, increased serum half-life, increased resistance to inhibitors, reduced renal clearance, reduced immunogenicity and/or increased bioavailability. The conjugate offers a number of advantages over the currently available APC products, including longer duration between injections, administration of less protein, and fewer side effects. Moreover, a reduced anticoagulant activity is beneficial to reduce the risk of bleeding while maintaining the antiinflammatory activity of APC (activated protein C) conjugates. This must be especially important when the conjugate has an extended plasma life. The gene for protein C is located on chromosome 2q13-q14. The present sequence encodes precursor protein C.

Sequence 1383 BP; 286 A; 418 C; 440 G; 239 T; 0 other;

#### Alignment Scores:

Pred. No.:	5	08e-20	Length:	1383
Score:	160.00	Matches:	31	
Percent Similarity:	75.008	Conservative:	2	
Best Local Similarity:	70.458	Mismatches:	11	
Query Match:	81.228	Indels:	0	
DB:	24	Gaps:	0	

SEQ1-4EDITS (1-44) x ABK86038 (1-1383)

OY	1	AlaAsnSerPheLeu*****LeuArgGlnGlySerLeu**Arg**CysIle*****	20
DB	127	CCCAACTCTCTCTCGAGCAGCTCCCTCACAGCAGCCGCGAGTCATAGAGAG	186
OY	21	IleCysAspPhe*****AlaLys***IlePheGluAspValaspsprhIleuAlaI	40
DB	187	ATCTGTGACTTCGAGGAGCCAGGAATTTTCAAATGTGATGACACACTGCGCTTC	246
OY	41	TrpSerLysHis 44	
DB	247	TGCTCCAAGCAC 258	

Search completed: May 23, 2003, 06:10:27  
Job time : 251 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.4-p5-4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: May 23, 2003, 06:03:42 ; Search time 1761 Seconds

(without alignments)  
404.657 Million cell updates/sec

Title: SEQ1-4EDITS

Perfect score: 197

Sequence: 1 ANSLFLXLRGSLXKRCIX.....XXAKXIFEDVDTLAFWSKH 44

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08

Maximum Match 1008

Listing first 45 summaries

Command line parameters: -DEV-xlh

-MODEL=frame.p2n.model -DEV-xlh

-O=seq2.1/USPNC-SPOOL/SCNIZER51/runat\_16052003\_160655\_26030/app\_query.fasta.1.199

-DB=EST -OFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45

-DOCCALIGN=200 -THR.SCORE=dot -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL

-OUTFMT=plc -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USER=SCNIZER51.BCGN.1.1906.0runat.16052003.160655.26030 -NCPU=6 -ICPU=3

-NO\_XIPXY -NO\_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONLOG -DEV.TIMEOUT=120

-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7

-TCAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estlin:\*

4: em\_estlmu:\*

5: em\_estlov:\*

6: em\_estlpl:\*

7: em\_estpro:\*

8: em\_hic:\*

9: gb\_estcl:\*

10: gb\_estcl2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estclun:\*

16: em\_estclom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_luv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vtl:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_tod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	160	81.2	772	13	B1762861
2	160	81.2	876	9	AL531077
3	140	71.1	338	9	AA880625
4	140	71.1	418	9	AI036235
5	140	71.1	498	12	BF236813
6	140	71.1	540	9	AA245955
7	140	71.1	550	9	AA771326
8	140	71.1	556	9	AA270181
9	140	71.1	608	13	B1147512
10	140	71.1	651	12	BF532364
11	140	71.1	669	13	BC969623
12	140	71.1	692	13	B1328022
13	140	71.1	695	13	B1218251
14	140	71.1	699	13	B1554535
15	140	71.1	712	13	B1219238
16	140	71.1	741	13	B1331957
17	140	71.1	761	13	B1102396
18	140	71.1	767	12	BF531774
19	140	71.1	767	12	BF234618
20	140	71.1	767	13	B1146467
21	140	71.1	767	13	B1220070
22	140	71.1	772	13	B1247252
23	140	71.1	778	13	BC972357
24	140	71.1	780	13	B1247353
25	140	71.1	789	13	B1328943
26	140	71.1	790	9	AI098831
27	140	71.1	801	9	AI097856
28	140	71.1	816	12	BF531859
29	140	71.1	823	13	BC972358
30	140	71.1	847	13	B1217283
31	140	71.1	848	13	B1147686
32	140	71.1	854	13	BC971793
33	140	71.1	862	13	B1218479
34	140	71.1	871	13	B1328283
35	140	71.1	956	13	B1246532
36	134	68.0	771	9	AA986009
37	133.5	67.8	540	10	AV689626
38	133.5	67.8	608	10	AV698460
39	129	65.5	532	9	BE664869
40	127	64.5	617	9	AI286922
41	124	62.9	951	12	BF788285
42	123	62.4	515	13	B1342627
43	123	62.4	541	10	BE015120
44	123	62.4	565	12	BC834324
45	122	61.9	419	10	BE032707

## ALIGNMENTS

RESULT 1										
B1762861	B1762861	772 bp	mRNA	linear	EST 25-SEP-2001					
LOCUS	603048383F1 NIH_MGC_116	Human sapiens	CDNA clone	IMAGE:5188604 5'						
DEFINITION	mRNA sequence.									
ACCESSION	B1762861									
VERSION	B1762861.1	GI:15754427								
KEYWORDS	EST.									
SOURCE	human.									
ORGANISM	Homo sapiens									
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.									
	1 (bases 1 to 772)									
REFERENCE	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .									
AUTHORS	National Institutes of Health, Mammalian Gene Collection (MGC)									
TITLE										

JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-r@mail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINTL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINTL at:  
<http://image.llnl.gov>  
 Plate: L1AM1471 row: e column: 21  
 High quality sequence stop: 764.

FEATURES  
 source Location/Qualifiers  
 1.772  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:518604"  
 /clone\_1lb="NIH\_MGC\_116"  
 /lab\_host="DH10B"  
 /note="Organ: pooled colon, kidney, stomach; Vector:  
 pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA  
 source anonymous pool of 3 colons, age 26 yo male, 49 yo  
 female, 71 yo male colon; 46 yo male kidney, and pool of 2  
 stomachs, 62 yo male and 70 yo female. Library is  
 oligo-dT primed and directionally cloned (EcoRV site is  
 destroyed upon cloning). Average insert size 1.4 kb,  
 insert size range 1-3 kb. Library is normalized and  
 enriched for full-length clones and was constructed by C.  
 Gruber (Invitrogen). Research Genetics tracking code  
 023. Note: this is a NIH\_MGC Library."

BASE COUNT 157 a 228 c 248 g 139 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,95e-18 Length: 772  
 Score: 160.00 Matches: 31  
 Percent Similarity: 75.008 Conservative: 2  
 Best Local Similarity: 70.458 Mismatches: 11  
 Query Match: 81.228 Indels: 0  
 DB: 13 Gaps: 0

SEQ1-4EDITS (1-44) x B1762861 (1-772)

QY 1 AlaAsnSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 Db 183 GCCAAGCTCCTCTCGAGAGAGCGCTCGACAGAGCGCTGAGCGAGTCATAGAGAG 242

QY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspTrpLeuAlaPhe 40  
 Db 243 ATCTGTACTTCGAGAGGCGCAAGAAATTTTCCAAATGTGATGACACACCTGGCCTTC 302

QY 41 TrpSerLysHis 44  
 Db 303 TGGTCCACAGCAC 314

RESULT 2  
 LOCUS AL531077 876 bp mRNA linear EST 13-FEB-2001  
 DEFINITION AL531077 LTI\_NFL001\_NBC4 Homo sapiens cDNA clone CS0DM001Y109 5  
 prime, mRNA sequence.  
 ACCESSION AL531077  
 VERSION AL531077.1 GI:12794570  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 876)  
 AUTHORS Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - France  
 Email: [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr).

FEATURES  
 source Location/Qualifiers  
 1.876  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0DM001Y109"  
 /clone\_1lb="LTI\_NFL001\_NBC4"  
 /sex="male"  
 /tissue.type="neuroblastoma cells"  
 /lab\_host="DH10B"  
 /note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-stranded cDNA was digested with Not I and  
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
 vector. Library was normalized. Library was constructed  
 by Life Technologies. Contact: Feng Liang Life  
 Technologies, a division of Invitrogen 9800 Medical Center  
 Drive Rockville, Maryland 20850, USA Fax : (1) 301 610  
 8371 Email : [fliang@lifestech.com](mailto:fliang@lifestech.com) URL :  
<http://fulllength.invitrogen.com>"

BASE COUNT 166 a 256 c 303 g 150 t 1 others  
 ORIGIN

Alignment Scores:  
 Pred. No.: 2.32e-18 Length: 876  
 Score: 160.00 Matches: 31  
 Percent Similarity: 75.008 Conservative: 2  
 Best Local Similarity: 70.458 Mismatches: 11  
 Query Match: 81.228 Indels: 0  
 DB: 9 Gaps: 0

SEQ1-4EDITS (1-44) x AL531077 (1-876)

QY 1 AlaAsnSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 Db 180 GCCAAGCTCCTCTCGAGAGAGCGCTCGACAGAGCGCTGAGCGAGTCATAGAGAG 239

QY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspTrpLeuAlaPhe 40  
 Db 240 ATCTGTACTTCGAGAGGCGCAAGAAATTTTCCAAATGTGATGACACACCTGGCCTTC 299

QY 41 TrpSerLysHis 44  
 Db 300 TGGTCCACAGCAC 311

RESULT 3  
 LOCUS AA880625 338 bp mRNA linear EST 26-MAR-1998  
 DEFINITION vx41e03.r1 Stratagene mouse lung 937302 Mus musculus cDNA clone  
 IMAGE:1277788 5' similar to gb:X02750.cds3 PROTEIN C PRECURSOR  
 (HUMAN); gb:D10445 Mouse mRNA for protein C, complete cds (mouse);,  
 mRNA sequence.  
 ACCESSION AA880625  
 VERSION AA880625.1 GI:2989608  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 338)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, R., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 TITLE The WashU-HMIT Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HMIT Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800

Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:669588  
 Seq primer: -28m13 rev1 ET from Amerham  
 High quality sequence stop: 262.

## FEATURES

source

1.338

/organism="Mus musculus"

/strain="C57BL/6 x CBA"

/db\_xref="taxon:10090"

/clone\_1b="IMAGE:1277788"

/clone\_1lb="Stratagene mouse lung 937302"

/sex="female"

/tissue\_type="lung"

/dev\_stage="6-8 month old"

/lab\_host="SOLR (kanamycin resistant)"

/note="Organ: lung; Vector: pBluescript SK-; Site\_1: EcoRI

; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo

dt. 6-8 month old female lung and 1.5 year old male lung

were source of mRNA. Average insert size: 1.5 kb. Uni-ZAP

XR Vector: -5' adaptor sequence: 5' GAATTCGCCACGAG 3' -3'

adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' -3'

## BASE COUNT

72 a 98 c 96 g 72 t

## ORIGIN

## Alignment Scores:

pred. No.: 2.58e-15 Length: 338  
 Score: 140.00 Matches: 26  
 Percent Similarity: 75.00% Conservative: 7  
 Best Local Similarity: 59.09% Mismatches: 11  
 Query Match: 71.07% Indels: 0  
 DB: 9 Gaps: 0

SEQ1-4EDITS (1-44) x AA880625 (1-338)

QY 1 AlaasSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*CysIle\*\*\*\*\* 20  
 |||||  
 Db 163 GCCACAGCTCTCTGGAGAGATGCGGCCGACGCTGACGGAGCTATGAGAGAG 222

QY 21 IlcCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
 |||||  
 Db 223 ATCTGTGACTTCGAGAGAGCCGAGAGATTTCCAAAATGTGGAGACACACTGCGCTTC 282

QY 41 TrpSerIysHis 44  
 |||||

Db 283 TGGATCAAGTAC 294

## RESULT 4

AI036235

LOCUS 418 bp mRNA linear EST 26-JUN-1998  
 DEFINITION v183c03.r1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA

clone IMAGE:987268 5' similar to gb:DI0445 Mouse mRNA for protein  
 C, complete cds (MOUSE);, mRNA sequence.

ACCESSION AI036235  
 VERSION AI036235.1 GI:325926

KEYWORDS EST.  
 SOURCE house mouse.

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 418)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Watsling,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

TITLE The WashU-HIMI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT Contact: Marra M/Mouse EST Project  
 WashU-HIMI Mouse EST Project  
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@watson.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 MGI:559548  
 Seq primer: -28m13 rev1 ET from Amerham  
 High quality sequence stop: 214.

## FEATURES

source

1.418

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone\_1b="IMAGE:987268"

/clone\_1lb="Stratagene mouse diaphragm (#937303)"

/tissue\_type="diaphragm"

/dev\_stage="adult"

/lab\_host="SOLR (kanamycin resistant)"

/note="Organ: diaphragm; Vector: pBluescript SK-; Site\_1:

EcoRI; Site\_2: XhoI; Cloned unidirectionally from mRNA

prepared from diaphragm muscle. Primer: Oligo dt. Average

insert size: 1.5 kb. Uni-ZAP XR Vector: -5' adaptor

sequence: 5' GAATTCGCCACGAG 3' -3' adaptor sequence: 5'

CTCGAGTTTCTTTTCTTTT 3' -3'

## BASE COUNT

86 a 117 c 125 g 89 t 1 others

## ORIGIN

## Alignment Scores:

pred. No.: 3.46e-15 Length: 418  
 Score: 140.00 Matches: 26  
 Percent Similarity: 75.00% Conservative: 7  
 Best Local Similarity: 59.09% Mismatches: 11  
 Query Match: 71.07% Indels: 0  
 DB: 9 Gaps: 0

SEQ1-4EDITS (1-44) x AI036235 (1-418)

QY 1 AlaasSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*CysIle\*\*\*\*\* 20  
 |||||  
 Db 168 GCCACAGCTCTCTGGAGAGATGCGGCCGACGCTGACGGAGCTATGAGAGAG 227

QY 21 IlcCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
 |||||  
 Db 228 ATCTGTGACTTCGAGAGAGCCGAGAGATTTCCAAAATGTGGAGACACACTGCGCTTC 287

QY 41 TrpSerIysHis 44  
 |||||

Db 288 TGGATCAAGTAC 299

## RESULT 5

BF236813

LOCUS 498 bp mRNA linear EST 14-NOV-2000  
 DEFINITION 602027916F1 NCI-CGAP-L19 Mus musculus cDNA clone IMAGE:4163186 5'

mRNA sequence.  
 ACCESSION BF236813  
 VERSION BF236813.1 GI:11150730

KEYWORDS EST.  
 SOURCE house mouse.

## ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 498)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov

Tissue Procurement: Jeffrey E. Green, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Incyte Genomics, Inc.

CDNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov  
 Plate: L1AM9446 row: p column: 03  
 High quality sequence stop: 498.  
 Location/Qualifiers

## FEATURES

SOURCE

1. .498  
 /organism="Mus musculus"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:4163186"  
 /clone\_id="NCI CGAP\_L19"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: Liver; Vector: PCMV-Spore; Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dt.  
 Average insert size 1.9 kb. Constructed by Life  
 Technologies. Note: this is a NCI CGAP Library."  
 BASE COUNT 107 a 140 c 147 g 104 t

## ALIGNMENT SCORES:

Pred. No.: 4.41e-15 Length: 498  
 Score: 140.00 Matches: 26  
 Percent Similarity: 75.00% Conservative: 7  
 Best Local Similarity: 59.09% Mismatches: 11  
 Query Match: 71.07% Indels: 0  
 Gaps: 0

SEQ1-4EDITS (1-44) x BF236813 (1-498)

QY 1 Alaansserphleu\*\*\*\*\*LeuArgInGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 |||||  
 Db 218 GCCACAGCTCTCGAGAGATGCGCGCAGCGCAGCTGTATGAGAGAG 277  
 |||||  
 QY 21 Illecysasphe\*\*\*\*\*AlaLys\*\*\*IllephgluaspValaspPThrLeuAlaPhe 40  
 |||||  
 Db 278 ATCTGTACTTCGAGAGAGCCCGAGAGATTTCCTCAAAATGTGAGACACACTGGCCTTC 337  
 |||||  
 QY 41 Trpserlyshs 44  
 |||||  
 Db 338 TGGATCAAGTAC 349

## RESULT 6

AA245955

## LOCUS

AA245955 540 bp mRNA linear EST 10-MAR-1997  
 m02040.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:678990 5'

## DEFINITION

similar to gb:DI0445 Mouse mRNA for protein C, complete cds (MOUSE  
 ); mRNA sequence.

## ACCESSION

AA245955  
 AA245955.1 GI:1876751

## VERSION

AA245955

## KEYWORDS

EST.

## SOURCE

house mouse.  
 Mus musculus

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

## REFERENCE

1 (bases 1 to 540)

## AUTHORS

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

## TITLE

The Washu-HMI Mouse EST Project

## JOURNAL

Unpublished (1996).

## COMMENT

Contact: Marra M/Mouse EST Project  
 Washu-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:418694  
 Seq primer: -28m13 rev2 ET from Amersham  
 High quality sequence stop: 352.

## FEATURES

SOURCE

Location/Qualifiers  
 1. .540  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:678990"  
 /clone\_id="Soares mouse NML"  
 /tissue\_type="Liver"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker. Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA  
 was primed with a Not I - oligo(dt) primer [5'  
 TGTACCAATCTGAGAGAGCGCGCGCAATCTTTTCTTTTCTT 3'];  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 constructed and normalized by Bento Soares and M.Fatima  
 Bonaldo."  
 BASE COUNT 112 a 149 c 166 g 113 t

## ALIGNMENT SCORES:

Pred. No.: 4.94e-15 Length: 540  
 Score: 140.00 Matches: 26  
 Percent Similarity: 75.00% Conservative: 7  
 Best Local Similarity: 59.09% Mismatches: 11  
 Query Match: 71.07% Indels: 0  
 Gaps: 0

SEQ1-4EDITS (1-44) x AA245955 (1-540)

QY 1 Alaansserphleu\*\*\*\*\*LeuArgInGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 |||||  
 Db 186 GCCACAGCTCTCGAGAGATGCGCGCAGCGCAGCTGTATGAGAGAG 245  
 |||||  
 QY 21 Illecysasphe\*\*\*\*\*AlaLys\*\*\*IllephgluaspValaspPThrLeuAlaPhe 40  
 |||||  
 Db 246 ATCTGTACTTCGAGAGAGCCCGAGAGATTTCCTCAAAATGTGAGACACACTGGCCTTC 305  
 |||||  
 QY 41 Trpserlyshs 44  
 |||||  
 Db 306 TGGATCAAGTAC 317

## RESULT 7

AA771326

## LOCUS

AA771326 546 bp mRNA linear EST 29-JAN-1998  
 vm43b03.r1 Stragene mouse diaphragm (#937303) Mus musculus cDNA  
 clone IMAGE:1000973 5' similar to gb:DI0445 Mouse mRNA for protein  
 C, complete cds (MOUSE); mRNA sequence.

## ACCESSION

AA771326  
 AA771326.1 GI:2823137

## VERSION

AA771326

## KEYWORDS

EST.

## SOURCE

house mouse.  
 Mus musculus

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 546)

## AUTHORS

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,  
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,  
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,  
 Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and  
 Waterston,R.

## TITLE

The Washu-HMI Mouse EST Project

## JOURNAL

Unpublished (1996).

## COMMENT

Contact: Marra M/Mouse EST Project  
 Washu-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: mouseest@wustl.edu  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 MGI:565189









LOCUS	BI554535	699 bp	MRNA	linear	EST 05-SEP-2001
DEFINITION	603235665P1 NCL_CGAP_L19 Mus musculus cDNA clone IMAGE:5309848 5',				
ACCESSION	B1554535				
VERSION	B1554535.1 GI:15441849				
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 699)				
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .				
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)				
JOURNAL	Unpublished (1999)				
COMMENT	Contact: Robert Strusberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: Jeffrey E. Green, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: L19M11785 row: a column: 17 High quality sequence stop: 699. Location/Qualifiers 1..699 /organism="Mus musculus" /strain="FVB/N" /db_xref="taxon:10090" /cd_xref="IMAGE:5309848" /clone_lib="NCL_CGAP_L19" /lab_host="DH10B (T1 phage-resistant)" /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.9 kb. Constructed by Life Technologies. Note: this is a NCL_CGAP Library."				
BASE COUNT	162 a 184 c 214 g 139 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	7.06e-15	Length:	699		
Score:	140.00	Matches:	26		
Percent Similarity:	75.09%	Conservative:	7		
Best Local Similarity:	59.09%	Mismatches:	11		
Query Match:	71.07%	Indels:	0		
DB:	13	Gaps:	0		
Seq1-4EDITS (1-44) x B1554535 (1-699)					
Oy	1	AlaasnSerPheIeu*****LeuArgGlnGlySerIeu**Arg**CysIle*****	20		
Db	186	GCACACACCTTCCTCTGGAAGAGATGGCGCCAGGACCTGGAACGGGAGTGTATGGAGGAG	245		
Oy	21	IleCysAspPhe*****AlaIys***IlePheGluAspValAspAspThrIeuAlaPhe	40		
Db	246	ATCTGTGACTTCGAGAGGCCCGAGGAGATTTTCCAAATGTGGAAAGACACACTGCCTTC	305		
Oy	41	TrpSerLysSHs 44			
Db	306	TGGATCAAGTAC 317			
RESULT 15					
LOCUS	B1219238	712 bp	MRNA	linear	EST 11-JUL-2001
DEFINITION	602935956F1 NCL_CGAP_L19 Mus musculus cDNA clone IMAGE:5099389 5',				
ACCESSION	B1219238				
VERSION	B1219238.1 GI:14672682				
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				

```

REFERENCE          Eukaryotes: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
AUTHORS            Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Mus
TITLE              1 (bases 1 to 712)
JOURNAL            NIH-MGC http://mgc.nci.nih.gov/.
COMMENT            National Institutes of Health, Mammalian Gene Collection (MGC)
                   Unpublished (1999)
                   Contact: Robert Strausberg, Ph.D.
                   Email: cgaps-remail.nih.gov
                   Tissue Procurement: Jeffrey E. Green, M.D.
                   cDNA Library Preparation: Life Technologies, Inc.
                   cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
                   DNA Sequencing by: Incyte Genomics, Inc.
                   Clone distribution: MGC clone distribution information can be
                   found through the I.M.A.G.E. Consortium/LNLN at:
                   http://image.lnl.gov
                   Plate: LHAM1238 row: p column: 14
                   High quality sequence start: 2
                   High quality sequence stop: 609.
FEATURES           Location/Qualifiers
SOURCE             1..712
                   /organism="Mus musculus"
                   /strain="FVB/N"
                   /db_xref="taxon:10090"
                   /clone="IMAGE:509389"
                   /clone_lib="NCI CGAP Lib"
                   /lab_host="DH10B (T1 phage-resistant)"
                   /note="Organ: liver; Vector: pCMV-Sport6; Site_1: NCI;
                   Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                   Average insert size 1.9 kb. Constructed by Life
                   Technologies. Note: this is a NCI-CGAP Library."
BASE COUNT         167 a 186 c 222 g 137 t
ORIGIN
Alignment Scores:
Pred. NO.:          7.24e-15          Length:          712
Score:              140.00            Matches:          26
Percent Similarity: 75.00%            Conservative:     7
Best local Similarity: 59.09%          Mismatches:       11
Query Match:        71.07%            Indels:           0
DB:                 13                Gaps:            0
SEQ1-4EDITS (1-44) x B1219238 (1-712)
OY      1  AlasnsarPheleu*****LeuArgInGInGlySerLeu***Arg***CysIle***** 20
         |||||||
Db       71 GCCAACAGCTTCCTCGGAGAGATGCGGCCGACGAGCCTGGAACGGGAGTGTATGGAGCAG 130
OY      21 IleCysasphe*****AlaIys***IlePheGluaspValaspAspThrleuAlaphe 40
         |||||||
Db      131 ATCTGTGACTTCGAGGAGGCCGACGAGGATTTTCCAAATGTGGAAAGACACACTGGCCTTC 190
OY      41 TrpSerlyshs 44
         |||
Db      191 TGGATCAAGTAC 202

```

GenCore version 5.1.4-p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - nucleic search, using frame\_plus.p2n model

Run on: May 23, 2003, 06:06:29 ; Search time 54 Seconds  
(without alignments)  
249.885 Million cell updates/sec

Title: SEQ1-4EDITS  
Perfect score: 197  
Sequence: 1 ANSFLXXLRGSLXRCIXX.....XXAKXIFedVDTLAFMSKH 44

Scoring table: BLOSUM62  
Xgapop 10.0 , Ygapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2.1/USPTO/spool/SCHNIZER91/rnatc\_16052003\_160656\_26089/app\_query.fasta\_1.199  
-DB-Issued.Patents.NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsm62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pcio -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=SCHNIZER91 @CGN.1.1.40 @rnatc\_16052003\_160656\_26089 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAR -LARGEQUERY -NEG\_SCORES=0 -WAIT LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6 -FCGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

- 1: Issued.Patents.NA:\*
- 2: /cgn2.6/ptodata/1/lna/5A.COMB.seq:\*
- 3: /cgn2.6/ptodata/1/lna/5B.COMB.seq:\*
- 4: /cgn2.6/ptodata/1/lna/6A.COMB.seq:\*
- 5: /cgn2.6/ptodata/1/lna/6B.COMB.seq:\*
- 6: /cgn2.6/ptodata/1/lna/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	160	81.2	1386	2	US-08-756-506-3
2	160	81.2	1755	6	Sequence 3, Appl1 Patent No. 5225537-1
3	155	78.7	1387	6	Patent No. 5270178-1
4	119	60.4	11725	2	Sequence 1, Appl1 Patent No. 5270178
5	114	57.9	1554	1	Sequence 1, Appl1
6	114	57.9	1554	1	Sequence 1, Appl1
7	110	55.8	1404	4	Sequence 15, Appl1
8	110	55.8	1500	1	Sequence 4, Appl1
9	99	50.3	1440	1	Sequence 3, Appl1
10	99	50.3	1440	1	Sequence 3, Appl1
11	99	50.3	1440	1	Sequence 3, Appl1
12	99	50.3	1440	3	Sequence 13, Appl1

13	99	50.3	1440	3	US-09-009-656-13	Sequence 13, Appl1
14	99	50.3	1440	5	PCT-US93-04493-3	Sequence 3, Appl1
15	99	50.3	2422	1	US-08-475-845-1	Sequence 1, Appl1
16	99	50.3	2422	1	US-08-327-690-1	Sequence 1, Appl1
17	99	50.3	2422	2	US-08-660-289-1	Sequence 1, Appl1
18	99	50.3	2422	2	US-08-537-807-1	Sequence 1, Appl1
19	99	50.3	2422	2	US-08-871-003-1	Sequence 1, Appl1
20	99	50.3	2422	2	US-08-464-233-1	Sequence 1, Appl1
21	99	50.3	2422	4	US-09-189-607-1	Sequence 1, Appl1
22	99	50.3	2422	4	US-09-378-907-1	Sequence 1, Appl1
23	99	50.3	2422	5	PCT-US94-05779-1	Sequence 1, Appl1
24	99	50.3	2462	2	US-08-479-733A-25	Sequence 25, Appl1
25	99	50.3	2462	3	US-08-487-427-25	Sequence 25, Appl1
26	99	50.3	2462	3	US-08-479-727A-25	Sequence 25, Appl1
27	99	50.3	2462	3	US-08-482-369A-25	Sequence 25, Appl1
28	99	50.3	2462	5	PCT-US95-07439-25	Sequence 25, Appl1
29	85	43.1	3284	6	5258288-3	Patent No. 5258288
30	85	43.1	3290	1	US-07-985-691-1	Sequence 1, Appl1
31	85	43.1	3290	1	US-08-436-804-1	Sequence 1, Appl1
32	85	43.1	3290	1	US-08-267-387-1	Sequence 1, Appl1
33	85	43.1	6811	4	US-08-651-472-67	Sequence 67, Appl1
34	85	43.1	6811	4	US-08-358-928-67	Sequence 67, Appl1
35	84	42.6	1869	3	US-08-952-967-7	Sequence 7, Appl1
36	84	42.6	1947	1	US-07-998-972A-2	Sequence 2, Appl1
37	84	42.6	1947	1	US-08-463-953-2	Sequence 2, Appl1
38	84	42.6	1947	1	US-08-462-261-2	Sequence 2, Appl1
39	84	42.6	1947	2	US-08-479-733A-24	Sequence 24, Appl1
40	84	42.6	1947	3	US-08-487-427-24	Sequence 24, Appl1
41	84	42.6	1947	3	US-08-479-727A-24	Sequence 24, Appl1
42	84	42.6	1947	3	US-08-482-369A-24	Sequence 24, Appl1
43	84	42.6	1947	5	PCT-US92-11357-2	Sequence 24, Appl1
44	84	42.6	1947	5	PCT-US95-07439-24	Sequence 24, Appl1
45	84	42.6	1988	1	US-07-750-080A-15	Sequence 15, Appl1

## ALIGNMENTS

RESULT 1  
US-08-756-506-3  
; Sequence 3, Application US/08756506  
; Patent No. 5905185  
; GENERAL INFORMATION:  
; APPLICANT: Garner, Ian R.  
; APPLICANT: Cottingham, Ian R.  
; APPLICANT: Temperley, Simon M.  
; APPLICANT: Foster, Donald C.  
; APPLICANT: Sprecher, Cindy A.  
; APPLICANT: Prunkard, Donna E.  
; TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Zymogenetics, Inc.  
; STREET: 1201 Eastlake Avenue East  
; CITY: Seattle  
; STATE: WA  
; COUNTRY: USA  
; ZIP: 98102  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/756,506  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sawislak, Deborah A  
; REGISTRATION NUMBER: 37,438  
; REFERENCE/DOCKET NUMBER: 95-28  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 206-442-6672

```
TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1386 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1380
; US-08-756-506-3

Alignment Scores:
Pred. No.: 8.3e-21 Length: 1386
Score: 160.00 Matches: 31
Percent Similarity: 75.008 Conservative: 2
Best Local Similarity: 70.458 Mismatches: 11
Query Match: 81.228 Indels: 0
DB: 2 Gaps: 0

SEQ1-4EDITS (1-44) x US-08-756-506-3 (1-1386)
QY 1 AlaasSerPheleu*****LeuArgGInglySerleu***Arg**CysIle***** 20
DB 127 GCCAAGCTCTCTCTCGAGAGAGCTCCGTCACAGCAGCGCTGAGCGGAGCATAGAGAG 186
QY 21 IleCysAspPhe*****AlaLys***IlephgIuAspValAspAspThrleuAlaIaphe 40
DB 187 ATCTGTGACTTTCAGAGAGGCCAAGAAATTTTCCAAATGTGATGATGACACACTGGCCTTC 246
QY 41 TrpSerLysHis 44
DB 247 TGTGTCCAGCAC 258

RESULT 2
522537-1
; Patent No. 522537
; APPLICANT: FOSTER, DONALD
; TITLE OF INVENTION: METHODS FOR PRODUCING HYBRID
; PHOSPHOLIPID-BINDING PROTEINS
; NUMBER OF SEQUENCES: 14
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/459,082
; FILING DATE: 29-DEC-1989
; SEQ ID NO: 1
; LENGTH: 1755
522537-1

Alignment Scores:
Pred. No.: 1.1e-20 Length: 1755
Score: 160.00 Matches: 31
Percent Similarity: 75.008 Conservative: 2
Best Local Similarity: 70.458 Mismatches: 11
Query Match: 81.228 Indels: 0
DB: 6 Gaps: 0

SEQ1-4EDITS (1-44) x 522537-1 (1-1755)
QY 1 AlaasSerPheleu*****LeuArgGInglySerleu***Arg**CysIle***** 20
DB 196 GCCAAGCTCTCTCTCGAGAGAGCTCCGTCACAGCAGCGCTGAGCGGAGCATAGAGAG 255
QY 21 IleCysAspPhe*****AlaLys***IlephgIuAspValAspAspThrleuAlaIaphe 40
DB 256 ATCTGTGACTTTCAGAGAGGCCAAGAAATTTTCCAAATGTGATGATGACACACTGGCCTTC 315
QY 41 TrpSerLysHis 44
DB 316 TGTGTCCAGCAC 327

RESULT 3
5270178-1
```

```
; Patent No. 5270178
; APPLICANT: GERLITZ, BRUCE E.; GRINNELL, BRIAN W.
; TITLE OF INVENTION: VECTORS AND COMPOUNDS FOR EXPRESSION OF
; ZMOGEN FORMS OF HUMAN PROTEIN C
; NUMBER OF SEQUENCES: 21
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/484,133
; FILING DATE: 23-FEB-1990
; SEQ ID NO: 1
; LENGTH: 1387
5270178-1

Alignment Scores:
Pred. No.: 7.28e-20 Length: 1387
Score: 155.00 Matches: 30
Percent Similarity: 72.738 Conservative: 2
Best Local Similarity: 68.188 Mismatches: 12
Query Match: 78.688 Indels: 0
DB: 6 Gaps: 0

SEQ1-4EDITS (1-44) x 5270178-1 (1-1387)
QY 1 AlaasSerPheleu*****LeuArgGInglySerleu***Arg**CysIle***** 20
DB 127 GCCAAGCTCTCTCTCGAGAGAGCTCCGTCACAGCAGCGCTGAGCGGAGCATAGAGAG 186
QY 21 IleCysAspPhe*****AlaLys***IlephgIuAspValAspAspThrleuAlaIaphe 40
DB 187 ATCTGTGACTTTCAGAGAGGCCAAGAAATTTTCCAAATGTGATGATGACACACTGGCCTTC 246
QY 41 TrpSerLysHis 44
DB 247 TGTGTCCAGCAC 258

RESULT 4
US-08-756-506-1
; Sequence 1, Application US/08756506
; Patent No. 5905185
; GENERAL INFORMATION:
; APPLICANT: Garner, Ian
; APPLICANT: Cottingham, Ian R.
; APPLICANT: Temperley, Simon M.
; APPLICANT: Foster, Donald C.
; APPLICANT: Sprecher, Cindy A.
; APPLICANT: Prunkard, Donna E.
; TITLE OF INVENTION: PROTEIN C PRODUCTION IN TRANSGENIC
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ZymoGenetics, Inc.
; STREET: 1201 Eastlake Avenue East
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/756,506
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sawislak, Deborah A
; REGISTRATION NUMBER: 37,438
; REFERENCE/DOCKET NUMBER: 95-28
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-442-6672
; TELEFAX: 206-442-6678
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
```

```

: LENGTH: 11725 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(3520..3630, 5093..5117, 5210..5347, 5450
: LOCATION: ..5584, 8253..8395, 9269..9386, 10516..11102)
US-08-756-506-1

Alignment Scores:
Pred. No.: 5.9e-12 Length: 11725
Score: 119.00 Matches: 24
Percent Similarity: 71.05% Conservative: 3
Best Local Similarity: 63.16% Mismatches: 11
Query Match: 60.41% Indels: 0
DB: 2 Gaps: 0

SEQ1-4EDITS (1-44) x US-08-756-506-1 (1-11725)

Oy 1 AlaAsnSerPheLeu*****LeuArGInGlySerLeu***Arg**CysIle***** 20
Db 3520 GCCAACTCTCTCTCTGAGAGAGCTCCGTACACAGAGCTGGAGCGGAGTGCATAGAGAG 3579

Oy 21 IleCysAspPhe*****AlaLys***IlePheGluAspValAspAspThrLeu 38
Db 3580 ACTGTGACTTGGAGAGAGCCCAAGAAATTTTCAAAATGTGATGACACAGTA 3633

RESULT 5
US-08-469-486-1
: Sequence 1, Application US/08469486
: Patent No. 5739281
: GENERAL INFORMATION:
: APPLICANT: Thoegeisen, Hans Christian
: APPLICANT: Hollet, Thor Las
: TITLE OF INVENTION: Improved method for the refolding of
: TITLE OF INVENTION: PROTEINS
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/469,486
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/192,060
: FILING DATE: February 4, 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Paul T. Clark
: REGISTRATION NUMBER: 30,162
: REFERENCE/DOCKET NUMBER: 06363/002001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617 542 3070
: TELEFAX: 617 542 8906
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1554 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
```

```

: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: YES
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Bos taurus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 76..1551
US-08-469-486-1

Alignment Scores:
Pred. No.: 4.52e-12 Length: 1554
Score: 114.00 Matches: 20
Percent Similarity: 63.64% Conservative: 8
Best Local Similarity: 45.45% Mismatches: 16
Query Match: 57.87% Indels: 0
DB: 1 Gaps: 0

SEQ1-4EDITS (1-44) x US-08-469-486-1 (1-1554)

Oy 1 AlaAsnSerPheLeu*****LeuArGInGlySerLeu***Arg**CysIle***** 20
Db 196 GCCAACTCTCTCTCTGAGAGAGTGAAGCAACCTGGAGCGAGTGCCTGAGAGAG 255

Oy 21 IleCysAspPhe*****AlaLys***IlePheGluAspValAspAspThrLeuAlaPhe 40
Db 256 GCCTGCTCTACTGAGAGAGCCCGCCGAGCTCTTGCAGAGCGCAGAGCAGCAGCATGATTC 315

Oy 41 TrpSerLysHis 44
Db 316 TGGAGTAATATAC 327

RESULT 6
US-08-469-658-1
: Sequence 1, Application US/08469658
: Patent No. 5917018
: GENERAL INFORMATION:
: APPLICANT: Thoegeisen, Hans Christian
: APPLICANT: Hollet, Thor Las
: TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOOLDING OF
: TITLE OF INVENTION: PROTEINS
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson P.C.
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/469,658
: FILING DATE: June 5, 1995
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/192,060
: FILING DATE: February 4, 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Paul T. Clark
: REGISTRATION NUMBER: 30,162
: REFERENCE/DOCKET NUMBER: 06363/002002
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617 542 5070
: TELEFAX: 617 542 8906
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1554 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
```

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1554 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: YES  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Bos taurus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 76..1551  
US-08-469-658-1

Alignment Scores:  
Pred. No.: 4,52e-12 Length: 1554  
Score: 114.00 Matches: 20  
Percent Similarity: 63.64% Conservative: 8  
Best Local Similarity: 45.45% Mismatches: 16  
Query Match: 57.87% Indels: 0  
DB: 2 Gaps: 0

SEQ1-4EDITS (1-44) x US-08-469-658-1 (1-1554)

QY 1 AlaasSerPheLeu\*\*\*\*\*LeuArgInGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
DB 196 GCCAATTCCTTTCTTGAAGAGATGAAGAAAGACCTCGAAGAGAGATGCGTGGAGAG 255  
QY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
DB 256 GCCTGCTCATACGAGAGAGCCCGCGAGGCTCTTGCAGAGCAGCAGACGATGAATTC 315  
QY 41 TrpSerLysHis 44  
DB 316 TGGATTAATATAC 327

RESULT 7

US-09-202-101-15  
Sequence 15, Application US/09202101  
Patent No. 6277618

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Recombinant blood-coagulation proteases

NUMBER OF SEQUENCES: 17

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30B (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/202,101

FILING DATE:

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:

LENGTH: 1404 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-09-202-101-15

Alignment Scores:  
Pred. No.: 2,27e-11 Length: 1404  
Score: 110.00 Matches: 19  
Percent Similarity: 63.64% Conservative: 9  
Best Local Similarity: 43.18% Mismatches: 16  
Query Match: 55.84% Indels: 0  
DB: 4 Gaps: 0

SEQ1-4EDITS (1-44) x US-09-202-101-15 (1-1404)

QY 1 AlaasSerPheLeu\*\*\*\*\*LeuArgInGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
DB 67 GCCAATTCCTTTCTTGAAGAGATGAAGAAAGACCTCGAAGAGAGATGCGTGGAGAG 126  
QY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
DB 127 ACCTGCTCATACGAGAGAGCCCGCGAGGCTCTTGCAGAGCAGCAGACGATGAATTC 186  
QY 41 TrpSerLysHis 44  
DB 187 TGGATTAATATAC 198

RESULT 8

US-08-487-037-4  
Sequence 4, Application US/08487037  
Patent No. 5755863

GENERAL INFORMATION:

APPLICANT: Wolf, David L.

TITLE OF INVENTION: RECOMBINANT AGENTS AFFECTING THROMBOSIS

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 2000 Pennsylvania Avenue, NW

CITY: Washington

STATE: DC

COUNTRY: USA

ZIP: 20006-1812

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,037

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Adler, Reid G.

REGISTRATION NUMBER: 30,988

REFERENCE/DOCKET NUMBER: 2803-0002.02

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 887-1500

TELEFAX: (202) 887-0763

TELEX: 90-4030

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1500 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-487-037-4

Alignment Scores:  
Pred. No.: 2,46e-11 Length: 1500  
Score: 110.00 Matches: 19  
Percent Similarity: 63.64% Conservative: 9  
Best Local Similarity: 43.18% Mismatches: 16  
Query Match: 55.84% Indels: 0  
DB: 1 Gaps: 0

SEQ1-4EDITS (1-44) x US-08-487-037-4 (1-1500)

QY 1 AlaasSerPheLeu\*\*\*\*\*LeuArgInGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
DB 153 GCCAATTCCTTTCTTGAAGAGATGAAGAAAGACCTCGAAGAGAGATGCGTGGAGAG 212  
QY 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
DB 213 ACCTGCTCATACGAGAGAGCCCGCGAGGCTCTTGCAGAGCAGCAGACGATGAATTC 272  
QY 41 TrpSerLysHis 44  
DB 273 TGGATTAATATAC 284

```

RESULT 9
US-07-882-202A-3
Sequence 3, Application US/07882202A
Patent No. 5374617
GENERAL INFORMATION:
APPLICANT: Morrissey, James H.
APPLICANT: Comp, Philip C.
TITLE OF INVENTION: Treatment of bleeding with modified
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESSES:
ADDRESS: Richards, Medlock & Andrews
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: Texas
COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/882,202A
FILING DATE: 13-MAY-1992
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hansen, Eugenia S.
REGISTRATION NUMBER: 31,966
REFERENCE/DOCKET NUMBER: OMR# B34290
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-939-4500
TELEFAX: 214-939-4600
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1440 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Blood
FEATURE:
NAME/KEY: CDS
LOCATION: 36..1433
OTHER INFORMATION: /note= "Coding portion of human
US-07-882-202A-3"
Alignment Scores:
Sred. NO.: 2.78e-09
Score: 99.00 Length: 1440
Percent Similarity: 58.54% Matches: 20
Best Local Similarity: 48.78% Mismatches: 17
Query Match: 50.25% Indels: 0
DB: 1 Gaps: 0
SEQ1-4EDITS (1-44) x US-07-882-202A-3 (1-1440)
Oy 1 AlaAaSaSrPheLeu*****LeuAaRgInLlySerLeu***Arg****CysIle***** 20
Db 216 GCCACGGCTTCGAGAGACCTCGCGCGGCTCCCTCGAGAGAGGAGTGCACAGAGAG 275
Oy 21 IleCysAspPhe*****AlaLys***IlePheCInuSpValAspAPThLeuAlaPhe 40
Db 276 CAGTCTCTCTTCGAGAGAGCCCGGAGAGATCTTCAAGACAGCGGAGAGACGACGACTGTTCC 335
Oy 41 Trp 41
|||

```

```

Db          336 TGG 338

RESULT 10
US-08-021-615A-3
: Sequence 3, Application US/08021615A
: Patent No. 5504064
: GENERAL INFORMATION:
: APPLICANT:.. Morrissey, James H.
: APPLICANT: Comp, Philip C.
: TITLE OF INVENTION: Treatment of Bleeding with Modified
: TITLE OF INVENTION: Tissue Factor in Combination with an Activator of
: TITLE OF INVENTION: FVII
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Richards, Medlock & Andrews
: STREET: 1201 Elm Street, Suite 4500
: CITY: Dallas
: STATE: Texas
: COUNTRY: US
: ZIP: 75270-2197
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/021,615A
: FILING DATE: 19-FEB-1993
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/882,202
: FILING DATE: 13-MAY-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Hansen, Eugenia S.
: REGISTRATION NUMBER: 31,966
: REFERENCE/DOCKET NUMBER: OMRF B34290CIP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 214-939-4500
: TELEFAX: 214-939-4600
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1440 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEetical: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: TISSUE TYPE: Blood
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 36..1433
: OTHER INFORMATION: /note="Coding portion of human
: OTHER INFORMATION: factor VII cDNA"
US-08-021-615A-3

Alignment Scores:
Pred. No.: 2,78e-09 Length: 1440
Score: 99.00 Matches: 20
Percent Similarity: 58.54% Conservative: 4
Best local Similarity: 48.78% Mismatches: 17
Query Match: 50.25% Indels: 0
DB: 1 Gaps: 0

SEQ1-4EDITS (1-44) x US-08-021-615A-3 (1-1440)
OY 1 AlasnsrPheLeu*****LeuAtfgGInGlySerLeu**Arg**CysIle***** 20
Db 216 GCCAACCGCTTCTCGAGAGAGAGCTGCGCGCGGCTCCCTCTGAGAGAGGAGTGCACAGAGAGAG 275
OY 21 IleCysAsPhe*****AlaIys***IlePheGluAsPValAsPAsPThrLeuAlaPhe 40

```

Db	276	CAGTGCCTCTTCGAGAGAGCCCGCGAGATCTTCAAGAGACCGGAGAGACGAACTGTTTC	335
QY	41	Trp	41
Db	336	TGG	338



QY 1 AlaAsnSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
 Db 216 GCCAACCGCTTCTCGAGAGACCTCGCCGCCGGGCTCCTCGAGAGAGCGGAGTCCACAGAGAGAG 275

QY 21 ILecysAspPhe\*\*\*\*\*AlaIys\*\*IlePheGlnAspValAspAspTrnLeuAlaPhe 40  
 Db 276 CAGTGTCTCTTCGAGAGAGCCCGGGAGATCTTCAGAGAGCCCGGAGAGAGCAGAACCTGTT 335

QY 41 Trp 41  
 Db 336 TGG 338

RESULT 13  
US-00-000

US-09-009-656-13  
; Sequence 13, Application US/09009656  
; Patent No. 6132730

```

1 GENERAL INFORMATION:
2 APPLICANT: Thorpe, Philip E.
3 APPLICANT: King, Steven W.
4 APPLICANT: Gao, Boning
5 TITLE OF INVENTION: COMBINED TISSUE FACTOR AND FACTOR VIIA
6 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR COAGULATION AND TUMOR
7 TITLE OF INVENTION: TREATMENT
8 NUMBER OF SEQUENCES: 27
9 CORRESPONDENCE ADDRESS:
10 ADDRESSEE: Arnold, White & Durkee
11 STREET: P.O. Box 4433
12 CITY: Houston
13 STATE: Texas
14 COUNTRY: USA
15 ZIP: 77210
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE: Floppy disk
18 COMPUTER: IBM PC compatible
19 OPERATING SYSTEM: PC-DOS/MS-DOS
20 SOFTWARE: Patentln Release #1.0, Version #1.30
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/09/009,656
23 FILING DATE: Concurrently Herewith
24 CLASSIFICATION:
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: US 60/042,427
27 FILING DATE: 27-MAR-1997
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER: US 60/036,205
30 FILING DATE: 27-JAN-1997
31 PRIOR APPLICATION DATA:
32 APPLICATION NUMBER: US 60/035,920
33 FILING DATE: 22-JAN-1997
34 ATTORNEY/AGENT INFORMATION:
35 NAME: Hildier, David W.
36 REGISTRATION NUMBER: 41,071
37 REFERENCE/DOCKET NUMBER: US-537
38 TELECOMMUNICATION INFORMATION:
39 TELEPHONE: 512/418-3000
40 TELEFAX: 512/474-7577
41 INFORMATION FOR SEQ ID NO: 13:
42 SEQUENCE CHARACTERISTICS:
43 LENGTH: 1440 base pairs
44 TYPE: nucleic acid
45 STRANDEDNESS: single
46 TOPOLOGY: linear
47 US-09-009-656-13

```

**Alignment Scores:**

Pred. No.:	2,78e+09	length:	1440
Score:	99.00	Matches:	20
Percent Similarity:	58.54%	Conservative:	4
Best Local Similarity:	48.78%	Mismatches:	17
Query Match:	50.25%	Indels:	0
DB:	3	Gaps:	0

SEQ1-4EDITS (1-44) X US-09-009-656-13 (1-1440)

[illegible]

RESULT 14

PCIT-US93-04493-3  
; Sequence 3, Application PC/TUS9304493  
: GENERAL INFORMATION:

1 APPLICANT: Morrissey, James H.  
2 APPLICANT: Comp, Phillip C.  
3 TITLE OF INVENTION: Truncated Tissue Factor and FvIIa or  
4 TITLE OF INVENTION: FvIIa Activator for Blood Coagulation  
5 NUMBER OF SEQUENCES: 4  
6 CORRESPONDENCE ADDRESSES:  
7 ADDRESSEE: Richards, Medlock & Andrews  
8 STREET: 1201 Elm Street, Suite 4500  
9 CITY: Dallas  
10 STATE: Texas  
11 COUNTRY: US  
12 ZIP: 75270-2197  
13 COMPUTER READABLE FORM:  
14 MEDIUM TYPE: Floppy disk  
15 COMPUTER: IBM PC compatible  
16 OPERATING SYSTEM: PC-DOS/MS-DOS  
17 SOFTWARE: Patentin Release #1.0, Version #1.25  
18 CURRENT APPLICATION DATA:  
19 APPLICATION NUMBER: PCT/US93/04493  
20 FILING DATE: 19930512  
21 CLASSIFICATION:  
22 PRIOR APPLICATION DATA:  
23 APPLICATION NUMBER: US 07/882202  
24 FILING DATE: 13-MAY-1992  
25 PRIOR APPLICATION DATA:  
26 APPLICATION NUMBER: US 08/021615  
27 FILING DATE: 19-FEB-1993  
28 ATTORNEY/AGENT INFORMATION:  
29 NAME: Trujillo, Doreen Y.  
30 REGISTRATION NUMBER: 35,719  
31 REFERENCE/DOCKET NUMBER: OMRF B34290C1PC/PCT  
32 TELECOMMUNICATION INFORMATION:  
33 TELEPHONE: 214-939-4500  
34 TELEFAX: 214-939-4600  
35 INFORMATION FOR SEQ ID NO: 3:  
36 SEQUENCE CHARACTERISTICS:  
37 LENGTH: 1440 base pairs  
38 TYPE: NUCLEIC ACID  
39 STRANDEDNESS: double  
40 TOPOLOGY: linear  
41 MOLECULE TYPE: cDNA  
42 HYPOTHETICAL: NO  
43 ANTI-SENSE: NO  
44 ORIGINAL SOURCE:  
45 ORGANISM: Homo sapiens  
46 TISSUE TYPE: Blood  
47 FEATURE:

## ; LOCATION:

```

OTHER INFORMATION: /product="Tissue Factor"
OTHER INFORMATION: /note="Coding portion of human factor VIII cDNA"
OTHER INFORMATION: /citation=(11)
PCT-US93-04493-3

```

Pred. No.:	2.78e-09	Length:	1440
------------	----------	---------	------



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 23, 2003, 06:45:52 ; Search time 140 Seconds  
(Without alignments)  
415.002 Million cell updates/sec

Title: SEQ1-4EDITS

Perfect score: 197  
Sequence: 1 ANSFLXXLRGSLXRCIXX.....XXAKXIFEDVDTLAFWSKH 44

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Fgapop 10.0 , Fgapext 0.5	
Delop 6.0 , Delext 7.0	

Searched: 828747 segs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```
-MODEL-framec_p2n_model -DEV-xlh
-Q/cgcn2_1/USPFO.spool/SCHNIZERS91/runatc_16052003_160658_26166/app_query.fasta_1.199
-DB-published.Applications_NA -OFMT-fasta -SUFFIX-p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS-bits -START=1 -MATRIX-blosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MILEN=0
-MAXLEN=2000000000 -USER=SCHNIZERS91.gcgn.1.1.135_4runatc_16052003_160658_26166
-NCPU=6 -ICPU=3 -NO_XLIPX -NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7
```

Database :

```
1: Published Applications_NA.*
2: /cgcn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
3: /cgcn2_6/ptodata/2/pubpna/PC1_NEM_PUB.seq.*
4: /cgcn2_6/ptodata/2/pubpna/US06_NEM_PUB.seq.*
5: /cgcn2_6/ptodata/2/pubpna/US07_NEM_PUB.seq.*
6: /cgcn2_6/ptodata/2/pubpna/PC1US_PUBCOMB.seq.*
7: /cgcn2_6/ptodata/2/pubpna/US08_NEM_PUB.seq.*
8: /cgcn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgcn2_6/ptodata/2/pubpna/US09_NEM_PUB.seq.*
10: /cgcn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgcn2_6/ptodata/2/pubpna/US10_NEM_PUB.seq.*
12: /cgcn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cgcn2_6/ptodata/2/pubpna/US60_NEM_PUB.seq.*
14: /cgcn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	179	90.9	1386 9	US-10-182-263-11 Sequence 11, Appl
2	179	90.9	1386 9	US-10-182-263-12 Sequence 12, Appl
3	174	88.3	1386 9	US-10-182-263-9 Sequence 9, Appl1
4	174	88.3	1386 9	US-10-182-263-10 Sequence 10, Appl

#### ALIGNMENTS

```
RESULT 1
US-10-182-263-11
: Sequence 11, Application US/10182263
: Publication No. US20030022354A1
: GENERAL INFORMATION:
: APPLICANT: Gerlitz, Bruce E
: APPLICANT: Jones, Bryan E
: TITLE OF INVENTION: PROTEIN C DERIVATIVES
: FILE REFERENCE: X-13611
: CURRENT APPLICATION NUMBER: US/10/182,263
: CURRENT FILING DATE: 2002-07-22
: PRIOR APPLICATION NUMBER: 60/181948
: PRIOR FILING DATE: 2002-02-11
: PRIOR APPLICATION NUMBER: 60/189199
: PRIOR FILING DATE: 2000-03-14
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 11
: LENGTH: 1386
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-182-263-11
Alignment Scores: 6,29e-26 Length: 1386
Pred. No.: 179.00 Matches: 35
Score: 79.55% Conservative: 0
Percent Similarity:
```

5	160	81.2	1257 9	US-09-978-917A-3	Sequence 3, Appl1
6	160	81.2	1260 9	US-10-182-263-7	Sequence 7, Appl1
7	160	81.2	1386 9	US-09-978-917A-1	Sequence 1, Appl1
8	160	81.2	1386 9	US-10-182-263-8	Sequence 8, Appl1
9	160	81.2	1386 9	US-09-880-107-3670	Sequence 3670, Ap
10	139	70.6	1543 10	US-09-917-800A-1575	Sequence 1575, Ap
11	129	65.5	356 10	US-09-960-352-2042	Sequence 2042, Ap
12	129	65.5	399 10	US-09-960-352-1089	Sequence 1089, Ap
13	114	57.9	351 10	US-09-960-352-1118	Sequence 1118, Ap
14	114	57.9	414 10	US-09-960-352-632	Sequence 632, Ap
15	107	54.3	266 10	US-09-867-701-9727	Sequence 9727, Ap
16	107	54.3	465 9	US-09-918-995-27923	Sequence 27923, A
17	102	51.8	483 9	US-09-918-995-8429	Sequence 8429, Ap
18	99	50.3	2462 10	US-09-964-824A-289	Sequence 289, Ap
19	99	50.3	2462 10	US-09-880-107-2251	Sequence 2251, Ap
20	99	50.3	6098 9	US-10-109-458-2	Sequence 2, Appl1
21	87	44.2	12850 9	US-10-017-122-1	Sequence 1, Appl1
22	85	43.1	497 10	US-09-864-864-178	Sequence 178, Ap
23	85	43.1	3294 10	US-09-864-864-268	Sequence 268, Ap
24	85	43.1	3344 10	US-09-964-824A-550	Sequence 550, Ap
25	85	43.1	3344 10	US-09-954-456-2146	Sequence 2146, Ap
26	85	43.1	3344 10	US-09-880-107-2260	Sequence 2260, Ap
27	84.5	42.9	678 9	US-09-759-130B-309	Sequence 309, Ap
28	84.5	42.9	678 9	US-10-189-133-39	Sequence 39, Appl
29	84.5	42.9	2498 9	US-09-759-130B-308	Sequence 308, Appl
30	84.5	42.9	2498 9	US-10-189-133-38	Sequence 38, Appl
31	84	42.6	417 9	US-09-918-995-7828	Sequence 7828, Ap
32	84	42.6	2000 12	US-10-044-090-331	Sequence 331, Ap
33	80	40.6	1413 10	US-09-884-901-2	Sequence 2, Appl1
34	80	40.6	1548 10	US-09-150-811-6	Sequence 6, Appl1
35	80	40.6	2792 10	US-09-118-748-1	Sequence 1, Appl1
36	80	40.6	2804 9	US-10-132-829-2	Sequence 2, Appl1
37	72	36.5	432 10	US-09-960-352-14612	Sequence 14612, A
38	72	36.5	337 10	US-09-960-352-11749	Sequence 11749, A
39	72	36.5	361 10	US-09-960-352-1535	Sequence 1535, Ap
40	72	36.5	370 10	US-09-960-352-13894	Sequence 13894, A
41	69	35.0	465 10	US-09-960-352-12285	Sequence 12285, A
42	69	35.0	390 10	US-09-960-352-11495	Sequence 11495, A
43	67.5	34.3	26928 9	US-10-020-141-7	Sequence 7, Appl1
44	67.5	34.3	26928 10	US-09-880-107-2278	Sequence 2278, Ap
45	67	34.0	488 9	US-09-918-995-31119	Sequence 31119, A

```
Best Local Similarity: 79.55%      Mismatches: 9
Query Match: 90.86%      Indels: 0
DB: 9      Gaps: 0

SEQ1-4EDITS (1-44) x US-10-182-263-11 (1-1386)

QY 1 AlaasSerPheLeu*****LeuArgGInGlySerLeu***Arg***CysIle***** 20
    |||||
Db 127 GCCAACTCCTCTCTCGAGAGGCTCCGTCAGAGGAGCGCTGAGCGGAGTCATAGAGAG 186
QY 21 IleCysAspPhe*****AlaIys***IlePheGluAspValAspAspThrLeuAlaPhe 40
    |||||
Db 187 ATCTGTGACTTGAGAGAGCGCAAGAAATTTCGAAGATGTGATGACACACTGGCCTTC 246
QY 41 TrpSerLysHis 44
    |||||
Db 247 TGGTCCAAGCAC 258

RESULT 2
US-10-182-263-12
; Sequence 12, Application US/10182263
; Publication No. US20030022354A1
; GENERAL INFORMATION:
; APPLICANT: Gerlitz, Bruce E
; APPLICANT: Jones, Bryan E
; APPLICANT: Grinnell, Brian W
; TITLE OF INVENTION: PROTEIN C DERIVATIVES
; FILE REFERENCE: X-13611
; CURRENT APPLICATION NUMBER: US/10/182,263
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 60/181948
; PRIOR FILING DATE: 2002-02-11
; PRIOR APPLICATION NUMBER: 60/181919
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 12
; LENGTH: 1386
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-182-263-12

Alignment Scores:
Pred. No.: 6,296-26      Length: 1386
Score: 179.00      Matches: 35
Percent Similarity: 79.55%      Conservative: 0
Best Local Similarity: 79.55%      Mismatches: 9
Query Match: 90.86%      Indels: 0
DB: 9      Gaps: 0

SEQ1-4EDITS (1-44) x US-10-182-263-12 (1-1386)

QY 1 AlaasSerPheLeu*****LeuArgGInGlySerLeu***Arg***CysIle***** 20
    |||||
Db 127 GCCAACTCCTCTCTCGAGAGGCTCCGTCAGAGGAGCGCTGAGCGGAGTCATAGAGAG 186
QY 21 IleCysAspPhe*****AlaIys***IlePheGluAspValAspAspThrLeuAlaPhe 40
    |||||
Db 187 ATCTGTGACTTGAGAGAGCGCAAGAAATTTCGAAGATGTGATGACACACTGGCCTTC 246
QY 41 TrpSerLysHis 44
    |||||
Db 247 TGGTCCAAGCAC 258

RESULT 3
US-10-182-263-9
; Sequence 9, Application US/10182263
; Publication No. US20030022354A1
; GENERAL INFORMATION:
; APPLICANT: Gerlitz, Bruce E
; APPLICANT: Jones, Bryan E
; APPLICANT: Grinnell, Brian W
; TITLE OF INVENTION: PROTEIN C DERIVATIVES
```

```
; FILE REFERENCE: X-13611
; CURRENT APPLICATION NUMBER: US/10/182,263
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 60/181948
; PRIOR FILING DATE: 2002-02-11
; PRIOR APPLICATION NUMBER: 60/181919
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 1386
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-182-263-9

Alignment Scores:
Pred. No.: 6,296-25      Length: 1386
Score: 174.00      Matches: 34
Percent Similarity: 77.27%      Conservative: 0
Best Local Similarity: 77.27%      Mismatches: 10
Query Match: 88.32%      Indels: 0
DB: 9      Gaps: 0

SEQ1-4EDITS (1-44) x US-10-182-263-9 (1-1386)

QY 1 AlaasSerPheLeu*****LeuArgGInGlySerLeu***Arg***CysIle***** 20
    |||||
Db 127 GCCAACTCCTCTCTCGAGAGGCTCCGTCAGAGGAGCGCTGAGCGGAGTCATAGAGAG 186
QY 21 IleCysAspPhe*****AlaIys***IlePheGluAspValAspAspThrLeuAlaPhe 40
    |||||
Db 187 ATCTGTGACTTGAGAGAGCGCAAGAAATTTCGAAGATGTGATGACACACTGGCCTTC 246
QY 41 TrpSerLysHis 44
    |||||
Db 247 TGGTCCAAGCAC 258

RESULT 4
US-10-182-263-10
; Sequence 10, Application US/10182263
; Publication No. US20030022354A1
; GENERAL INFORMATION:
; APPLICANT: Gerlitz, Bruce E
; APPLICANT: Jones, Bryan E
; APPLICANT: Grinnell, Brian W
; TITLE OF INVENTION: PROTEIN C DERIVATIVES
; FILE REFERENCE: X-13611
; CURRENT APPLICATION NUMBER: US/10/182,263
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 60/181948
; PRIOR FILING DATE: 2002-02-11
; PRIOR APPLICATION NUMBER: 60/181919
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 1386
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-182-263-10

Alignment Scores:
Pred. No.: 6,296-25      Length: 1386
Score: 174.00      Matches: 34
Percent Similarity: 77.27%      Conservative: 0
Best Local Similarity: 77.27%      Mismatches: 10
Query Match: 88.32%      Indels: 0
DB: 9      Gaps: 0

SEQ1-4EDITS (1-44) x US-10-182-263-10 (1-1386)

QY 1 AlaasSerPheLeu*****LeuArgGInGlySerLeu***Arg***CysIle***** 20
    |||||
```

```
Db 127 GCCAAGCTCTTCGTGAGAGAGCTCCGTCACGGGAGCTCGAGCGGAGTGCATAGAGAG 186
QY 21 IIECyAspPhe*****AlaIys***IlePheGluAspValAspAspThrLeuAlaPhe 40
Db 187 ATCTGTGACTTCGAGAGAGCGCCAAAGAAATTTTCGAAGATGTGATGACACACTGGCCTTC 246
QY 41 TrpSerLysHis 44
Db 247 TGGTCCAAAGCAC 258

RESULT 5
US-09-978-917A-3
; Sequence 3, Application US/09978917A
; Publication No. US20030027299A1
; GENERAL INFORMATION:
; APPLICANT: Maxygen Aps; Maxygen Holdings
; TITLE OF INVENTION: Protein C or activated protein C-like molecules
; FILE REFERENCE: 0219us310 - protein C
; CURRENT APPLICATION NUMBER: US/09/978,917A
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1257
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1257)
US-09-978-917A-3

Alignment Scores:
Pred. No.: 3,51e-22 Length: 1257
Score: 160.00 Matches: 31
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 70.45% Mismatches: 11
Query Match: 81.22% Indels: 0
DB: 9 Gaps: 0

SEQ1-4EDITS (1-44) x US-09-978-917A-3 (1-1257)
QY 1 AlaAsnSerPheLeu*****LeuArgGInGlySerLeu***Arg***CysIle***** 20
Db 1 GCCAAGCTCTTCGTGAGAGAGCTCCGTCACAGCAGCTCGAGCGGAGTGCATAGAGAG 60
QY 21 IIECyAspPhe*****AlaIys***IlePheGluAspValAspAspThrLeuAlaPhe 40
Db 61 ATCTGTGACTTCGAGAGAGCGCCAAAGAAATTTTCGAAGATGTGATGACACACTGGCCTTC 120
QY 41 TrpSerLysHis 44
Db 121 TGGTCCAAAGCAC 132

RESULT 6
US-10-182-263-7
; Sequence 7, Application US/10182263
; Publication No. US20030022354A1
; GENERAL INFORMATION:
; APPLICANT: Gerltz, Bruce E
; APPLICANT: Jones, Bryan E
; APPLICANT: Grinnell, Brian W
; TITLE OF INVENTION: PROTEIN C DERIVATIVES
; FILE REFERENCE: X-13611
; CURRENT APPLICATION NUMBER: US/10/182,263
; PRIOR FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: 60/181948
; PRIOR FILING DATE: 2002-02-11
; PRIOR APPLICATION NUMBER: 60/189199
; PRIOR FILING DATE: 2000-03-14
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 1260
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-182-263-7

Alignment Scores:
Pred. No.: 3,52e-22 Length: 1260
Score: 160.00 Matches: 31
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 70.45% Mismatches: 11
Query Match: 81.22% Indels: 0
DB: 9 Gaps: 0

SEQ1-4EDITS (1-44) x US-10-182-263-7 (1-1260)
QY 1 AlaAsnSerPheLeu*****LeuArgGInGlySerLeu***Arg***CysIle***** 20
Db 1 GCCAAGCTCTTCGTGAGAGAGCTCCGTCACAGCAGCTCGAGCGGAGTGCATAGAGAG 60
QY 21 IIECyAspPhe*****AlaIys***IlePheGluAspValAspAspThrLeuAlaPhe 40
Db 61 ATCTGTGACTTCGAGAGAGCGCCAAAGAAATTTTCGAAGATGTGATGACACACTGGCCTTC 120
QY 41 TrpSerLysHis 44
Db 121 TGGTCCAAAGCAC 132

RESULT 7
US-09-978-917A-1
; Sequence 1, Application US/09978917A
; Publication No. US20030027299A1
; GENERAL INFORMATION:
; APPLICANT: Maxygen Aps; Maxygen Holdings
; TITLE OF INVENTION: Protein C or activated protein C-like molecules
; FILE REFERENCE: 0219us310 - protein C
; CURRENT APPLICATION NUMBER: US/09/978,917A
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1383
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1383)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (1)...(126)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (127)..(1383)
US-09-978-917A-1

Alignment Scores:
Pred. No.: 3,97e-22 Length: 1383
Score: 160.00 Matches: 31
Percent Similarity: 75.00% Conservative: 2
Best Local Similarity: 70.45% Mismatches: 11
Query Match: 81.22% Indels: 0
DB: 9 Gaps: 0

SEQ1-4EDITS (1-44) x US-09-978-917A-1 (1-1383)
QY 1 AlaAsnSerPheLeu*****LeuArgGInGlySerLeu***Arg***CysIle***** 20
Db 127 GCCAAGCTCTTCGTGAGAGAGCTCCGTCACAGCAGCTCGAGCGGAGTGCATAGAGAG 186
QY 21 IIECyAspPhe*****AlaIys***IlePheGluAspValAspAspThrLeuAlaPhe 40
Db 187 ATCTGTGACTTCGAGAGAGCGCCAAAGAAATTTTCGAAGATGTGATGACACACTGGCCTTC 246
QY 41 TrpSerLysHis 44
```

Db 247 TGSTCCAGCAC 258

RESULT 8  
US-10-182-263-8  
; Sequence 8, Application US/10182263  
; Publication No. US20030022354A1  
; GENERAL INFORMATION:  
; APPLICANT: Gerlitz, Bruce E  
; APPLICANT: Jones, Bryan E  
; APPLICANT: Grinnell, Brian W  
; TITLE OF INVENTION: PROTEIN C DERIVATIVES  
; FILE REFERENCE: X-13611  
; CURRENT APPLICATION NUMBER: US/10/182,263  
; CURRENT FILING DATE: 2002-07-22  
; PRIOR APPLICATION NUMBER: 60/181948  
; PRIOR FILING DATE: 2002-02-11  
; PRIOR APPLICATION NUMBER: 60/189199  
; PRIOR FILING DATE: 2000-03-14  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 8  
; LENGTH: 1386  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-182-263-8

Alignment Scores:  
Pred. No.: 3,98e-22 Length: 1386  
Score: 160.00 Matches: 31  
Percent Similarity: 75.008 Conservative: 2  
Best Local Similarity: 70.458 Mismatches: 11  
Query Match: 81.228 Indels: 0  
Gaps: 0

DB: 9

SEQ1-4EDITS (1-44) x US-10-182-263-8 (1-1386)

QY 1 AlaAsnSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
; 127 GCCAACTCTTCTCGAGGAGCGCTCGTCACAGCAGCCTGGAGCGGAGTCATAGAGAG 186

Db 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
; 187 ATCTGTGACTTTCAGAGAGGCCAAGAAATTTTCAAATGTGATGACACACTGGCCTTC 246

QY 41 TrpSerIysHis 44  
; 247 TGSTCCAGCAC 258

Db 247 TGSTCCAGCAC 258

RESULT 9  
US-09-880-107-3670  
; Sequence 3670, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO  
; CURRENT APPLICATION NUMBER: US/09/880,107  
; CURRENT FILING DATE: 2001-06-14  
; PRIOR APPLICATION NUMBER: US 60/211,379  
; PRIOR FILING DATE: 2000-06-14  
; PRIOR APPLICATION NUMBER: US 60/237,054  
; PRIOR FILING DATE: 2000-10-02  
; NUMBER OF SEQ ID NOS: 3950  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3670  
; LENGTH: 1843  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 X02750

US-09-880-107-3670

Alignment Scores:  
Pred. No.: 5,72e-22 Length: 1843  
Score: 160.00 Matches: 31  
Percent Similarity: 75.008 Conservative: 2  
Best Local Similarity: 70.458 Mismatches: 11  
Query Match: 81.228 Indels: 0  
Gaps: 0

DB: 10

SEQ1-4EDITS (1-44) x US-09-880-107-3670 (1-1843)

QY 1 AlaAsnSerPheLeu\*\*\*\*\*LeuArgGlnGlySerLeu\*\*\*Arg\*\*\*CysIle\*\*\*\*\* 20  
; 224 GCCAACTCTTCTCGAGGAGCGCTCGTCACAGCAGCCTGGAGCGGAGTCATAGAGAG 283

Db 21 IleCysAspPhe\*\*\*\*\*AlaIys\*\*\*IlePheGluAspValAspAspThrLeuAlaPhe 40  
; 284 ATCTGTGACTTTCAGAGAGGCCAAGAAATTTTCAAATGTGATGACACACTGGCCTTC 343

QY 41 TrpSerIysHis 44  
; 344 TGSTCCAGCAC 355

Db 344 TGSTCCAGCAC 355

RESULT 10  
US-09-917-800A-1575  
; Sequence 1575, Application US/09917800A  
; Patent No. US20020119462A1  
; GENERAL INFORMATION:  
; APPLICANT: Mendrick, Donna  
; APPLICANT: Porter, Mark  
; APPLICANT: Johnson, Kory  
; APPLICANT: Castle, Arthur  
; APPLICANT: Elashoff, Michael  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Molecular Toxicology Modeling  
; FILE REFERENCE: 44921-5038-US  
; CURRENT APPLICATION NUMBER: US/09/917,800A  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,040  
; PRIOR FILING DATE: 2000-07-31  
; PRIOR APPLICATION NUMBER: US 60/222,880  
; PRIOR FILING DATE: 2000-11-02  
; PRIOR APPLICATION NUMBER: US 60/290,029  
; PRIOR FILING DATE: 2001-05-11  
; PRIOR APPLICATION NUMBER: US 60/290,645  
; PRIOR FILING DATE: 2001-05-15  
; PRIOR APPLICATION NUMBER: US 60/292,336  
; PRIOR FILING DATE: 2001-05-22  
; PRIOR APPLICATION NUMBER: US 60/295,798  
; PRIOR FILING DATE: 2001-06-06  
; PRIOR APPLICATION NUMBER: US 60/297,457  
; PRIOR FILING DATE: 2001-06-13  
; PRIOR APPLICATION NUMBER: US 60/298,884  
; PRIOR FILING DATE: 2001-06-19  
; PRIOR APPLICATION NUMBER: US 60/303,459  
; PRIOR FILING DATE: 2001-07-09  
; NUMBER OF SEQ ID NOS: 1740  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1575  
; LENGTH: 1543  
; TYPE: DNA  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 NM\_012803

US-09-917-800A-1575

Alignment Scores:  
Pred. No.: 7,25e-18 Length: 1543  
Score: 139.00 Matches: 26  
Percent Similarity: 75.008 Conservative: 7  
Best Local Similarity: 59.098 Mismatches: 11  
Query Match: 70.568 Indels: 0

```

DB:                               10          Gaps:                                0
SEQ1-4EDITS (1-44) x US-09-917-800A-1575 (1-1543)

OY      1 AlaanserPheLeu*****LeuArgGlnGlySerLeu***Arg**CysIle***** 20
         |||||||
         :|||   |||||   |||   |||
Db       172 GCCAACACCTTCCTCGAGAGAGTGGCGGCAGCACGCCGGAAGCTGTATGAGAGAG 231
         |||cysasrPhe*****AlaLys***IlePheGluaspValaspAspThrLeuAlaPhe 40
         |||||||
         |||:|   |||:|   |||:|   |||:|   |||:|   |||:|   |||:|   |||:|
Db       232 ATCTGTGACTTCGAGAGAGCCGACGAGATTTCACAAGATGTGAAACACACTGGCCTTT 291
         OY      41 TrpSerLysHis 44
         |||   |||:|   |||:|   |||:|   |||:|   |||:|   |||:|
Db       292 TGGATCAAGTAGC 303

RESULT 11
US-09-960-352-2042
: Sequence 2042, Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengbing
: APPLICANT: Byatt, John C.
: APPLICANT: Mathialagan, Nagappan
: TITLE OF INVENTION: MUSCLE ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
: FILE REFERENCE: 16511.006/37-21(10298)C
: CURRENT APPLICATION NUMBER: US/09/960,352
: CURRENT FILING DATE: 2001-09-24
: NUMBER OF SEQ ID NOS: 15112
: SEQ ID NO 2042
: LENGTH: 356
: TYPE: DNA
: ORGANISM: Bos taurus
: OTHER INFORMATION: Clone ID: 09-LTB34-047-Q1-E1-C1
US-09-960-352-2042

Alignment Scores:
Pred. NO.:                1,12e-16              Length:           356
Score:                    129.00                 Matches:            22
Percent Similarity:      72.73%                  Conservative:       10
Best Local Similarity:    50.00%                  Mismatches:         12
Query Match:              65.48%                  Indels:              0
DB:                        10                      Gaps:               0

SEQ1-4EDITS (1-44) x US-09-960-352-2042 (1-356)

OY      1 AlaanserPheLeu*****LeuArgGlnGlySerLeu***Arg**CysIle***** 20
         |||||||
         :|||   |||||   |||   |||
Db       183 GCCAACCTTCCTCGAGAGAGTGGCGGCAGCACGCCGGAAGCTGTATGAGAGAG 242
         |||cysasrPhe*****AlaLys***IlePheGluaspValaspAspThrLeuAlaPhe 40
         |||||||
         |||:|   |||:|   |||:|   |||:|   |||:|   |||:|   |||:|   |||:|
Db       243 GTCTGTGACTTCGAGAGAGCTGGGAGATTTCACAAAACAGAGACACATGGCCTTC 302
         OY      41 TrpSerLysHis 44
         |||||||
         |||   |||:|   |||:|   |||:|   |||:|   |||:|   |||:|
Db       303 TGGTCAAGTAGT 314

RESULT 12
US-09-960-352-1089
: Sequence 1089, Application US/09960352
: Patent No. US20020137139A1
: GENERAL INFORMATION:
: APPLICANT: Warren, Wesley C.
: APPLICANT: Tao, Nengbing
: APPLICANT: Byatt, John C.
: APPLICANT: Mathialagan, Nagappan
: TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
: TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
: FILE REFERENCE: 16511.006/37-21(10298)C
: CURRENT APPLICATION NUMBER: US/09/960,352

```

```

CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 1089
LENGTH: 399
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: unsure
LOCATION: (367)
OTHER INFORMATION: unsure at all n locations
OTHER INFORMATION: Clone ID: 05-LIB34-031-Q1-E1-B1
US-09-960-352-1089

Alignment Scores:
Pred. No.: 1.3e-16 Length: 399
Score: 129.00 Matches: 22
Percent Similarity: 72.73% Conservative: 10
Best Local Similarity: 50.00% Mismatches: 12
Query Match: 65.48% Indels: 0
DB: 10 Gaps: 0

SEQ1-4EDITS (1-44) x US-09-960-352-1089 (1-399)
QY 1 AlaasSerPheLeu*****LeuArgGlnGlySerLeu***Arg**CysIle***** 20
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 187 GCCAAGCTCTTCTGGAGAGCTCGCGCCGCGCAACGTGAGCGTGGAGCTCTCAGAGCAG 246
QY 21 IleCysAspPhe*****AlaLys**IlePheGluAspValAspAspThrLeuAlaPhe 40
    ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 247 GCTGTGAGTTTCGAGAGAGCTCGGAGATTTTCCAAACAGCAGACACATGCGCTTC 306
QY 41 TrpSerLysHis 44
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 307 TGGTCCAGGTAT 318

RESULT 13
US-09-960-352-1118
; Sequence 1118, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Byalt, John C.
; APPLICANT: Mathalaagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960.352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 1118
; LENGTH: 351
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 05-LIB34-061-Q1-E1-B1
US-09-960-352-1118

Alignment Scores:
Pred. No.: 1.1e-13 Length: 351
Score: 114.00 Matches: 20
Percent Similarity: 63.64% Conservative: 8
Best Local Similarity: 45.45% Mismatches: 16
Query Match: 57.87% Indels: 0
DB: 10 Gaps: 0

SEQ1-4EDITS (1-44) x US-09-960-352-1118 (1-351)
QY 1 AlaasSerPheLeu*****LeuArgGlnGlySerLeu***Arg**CysIle***** 20
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 137 GCCAAGCTCTTCTGGAGAGCTGAGCAGGAAACCTGAGAGCGAGAGTGGCTGAGAGAG 196
QY 21 IleCysAspPhe*****AlaLys**IlePheGluAspValAspAspThrLeuAlaPhe 40
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

D<sub>b</sub> 197 GCCTGCTCATTAGAGGAGGGCCCCGGAGGTCTTCAGAGGACCGAGCAGACGAGTGAATTC 250

QY 41 TrpSerLysHis 44  
|||||||:::  
D<sub>b</sub> 257 TGGAGTAAATTC 268

```

RESULT 14
US-09-960-352-632
: Sequence 632, Application US/09960352
: Patent No. US20020137139A1
:
GENERAL INFORMATION:
:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
:
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
:
FILE REFERENCE: 16511.006/37-21(10298)C
:
CURRENT APPLICATION NUMBER: US/09/960.352
:
CURRENT FILING DATE: 2001-09-24
:
NUMBER OF SEQ ID NOS: 15112
:
SEQ ID NO 632
:
LENGTH: 414
:
TYPE: DNA
:
ORGANISM: Bos taurus
:
OTHER INFORMATION: Clone ID: 03-LIB34-038-Q1-E1-A3
:
US-09-960-352-632

```

Alignment Scores:	
Pred. No.:	1,36e-13
Score:	114.00
Percent Similarity:	63.64%
Best Local Similarity:	45.45%
Query Match:	57.87%
DB:	10
Length:	414
Matches:	20
Conservative:	11
Mismatches:	16
Indels:	0
Gaps:	0

SEQ1-4EDITS (1-44) X US-09-960-352-632 (1-414)

[illegible]

Qy	41	TrpSerLyshis	44
			:
Db	256	TCGAGTAATAAC	267

RESULT 15  
US-09-867-701-9727  
; Sequence 9727, Application US/09867701  
; Patent No. US20020132237A1  
; Current: Incomplete

```

1  APPLICANT: Agiate, Paul A.
2  APPLICANT: Jones, Robert
3  APPLICANT: Harlocker, Susan L.
4  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
5  OF INFECTIONS AND DIAGNOSIS OF OVARIAN CANCER
6  FILE REFERENCE: 210121.497
7  CURRENT APPLICATION NUMBER: US/09/867,701
8  CURRENT FILING DATE: 2001-05-29
9  NUMBER OF SEQ ID NOS: 10912
10 SOFTWARE: fastseq for Windows Version 4.0
11 SEQ ID NO 9727
12     LENGTH: 266
13     TYPE: DNA
14 ORGANISM: Homo sapien
15 US-09-867-701-9727

```

Alignment Scores:  
Pred. No.: 1.95e-12 Length: 266

Score:	107.00	Matches:	18
Percent Similarity:	63.41%	Conservative:	8
Best Local Similarity:	43.90%	Mismatches:	15
Query Match:	54.31%	Indels:	0
DB:	10	Gaps:	0

SEQ1-4EDITS (1-44) X US-09-867-701-9727 (1-266)

```

QY      1 AlasanserPhleau*****LeuArgIIndIyserleu***Arg****CysIle***** 20
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      59 GCCAATAGATTCTCTCGAGAGAGACTGCCCGACGCCATCGAGCCAGAGTGCATGAGAGAG 118

QY      21 IleCysasphe*****AlaIys***IlephcIuAspValAspAspThrLeuAlaPhe 40
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      119 ATCTGCACCTACGAGGAGGTCACAGGAAGTGTTTTGAAGACCAAGAGAAACGATGGAAGTTC 178

QY      41 Trp 41
      ||||
Db      179 TGG 181

```

Search completed: May 23, 2003, 09:07:35  
Job time : 156 secs